Army Science Conference Proceedings

12-15 June 1990

Volume II

Principal Authors G through M

This document has been approved for public release and sale; its distribution is unlimited.

DTIC QUALITY INSPECTED 4

19970929 066

SARD-TR

30 JUL 1990

MEMORANDUM FOR SEE DISTRIBUTION

SUBJECT: Proceedings of the 1990 Army Science Conference

The 17th Army Science Conference was held at the Omni Durham Hotel and Convention Center, Durham, North Carolina, 12-15 June 1990. The conference presented a cross section of the many significant scientific and engineering programs carried out by the Department of the Army (DA). Additionally, it provided an opportunity for DA civilian and military scientists and engineers to present the results of their research and development efforts before a distinguished and critical audience.

These Proceedings of the the 1990 Army Science Conference are a compilation of all papers presented at the conference and the supplemental papers that were submitted.

Our purpose for soliciting these papers was to:

a. Stimulate the involvement of scientific and engineering talent within the Department of the Army;

b. Demonstrate Army competence in research and development;

c. Provide a forum wherein Army personnel can demonstrate the full scope and depth of their current projects; and

d. Promote the interchange of ideas among members of the Army scientific and engineering community.

The information contained in these volumes will be of benefit to those who attended the conference and to others interested in Army research and development. It is requested that these Proceedings be placed in technical libraries where they will be available for reference.

George T. Singley III
Deputy Assistant Secretary
For Research and Technology

DISTRIBUTION:
Office of the Under Secretary of Defense for Research and Advanced Technology, Wash, DC 20310-3080
Office of the Assistant Secretary of the Army (RD&A), ATTN: SARD-TR, Wash, DC 20310-0103
HQDA, Assistant Chief of Staff for Intelligence, ATTN: DAMI-ZA, Wash, DC 20310
HQDA, Deputy Chief of Staff for Logistics, ATTN: DALO-ZX, Wash, DC 20310-0500
HQDA, Deputy Chief of Staff for Operations & Plans, ATTN: DAMO-FD, Wash, DC 20310-0400
HQDA, Director of Information Systems for C4, ATTN: SAIS-ZA, Wash, DC 20310-0400
HQDA, Deputy Chief of Staff for Personnel, ATTN: DAPE-ZA, Wash, DC 20310-0300
Office of the Chief of Engineers, ATTN: CERD-M, 20 Massachusetts Avenue, N.W., Wash, DC 20314-1000
Office of the Surgeon General, ATTN: DASG-ZA, 5109 Leesburg Pike, Falls Church, VA 22041-3258
COMMANDERS/DIRECTORS:
USA Concepts Analysis Agency, 8120 Woodmont Avenue, Bethesda, MD 20814-2797
USA Operational Test & Evaluation Agency, 5600 Columbia Pike, Falls Church, VA 22041-5115
USA Information Systems Engineering Command, Ft. Huachuca, AZ 85613-5000
USA Strategic Defense Command, P. O. Box 15280, Arlington, VA 22215-0150
USA Materiel Command, ATTN: AMCLD, 5001 Eisenhower Avenue, Alexandria, VA 22333-0001
  USA Materiel Systems Analysis Activity, ATTN: AMXSY-D, APG, MD 21005-5071
  USA Armament, Munitions & Chemical Command, ATTN: AMSMC-CG, Rock Island, IL 61299-6000
  USA Armament RD&E Center, ATTN: SMCAR-TD, Picatinny Arsenal, NJ 07806-5000
    Fire Support Armament Cen, ATTN: SMCAR-FS, PA, NJ 07806-5000
    Armament Engineering Dir, ATTN: SMCAR-AE, PA, NJ 07806-5000
    Close Combat Armament Cen, ATTN: SMCAR-CC, PA, NJ 07806-5000
    Chemical RD&E Center, ATTN: SMCCR-TD, APG, MD 21010-5423
USA Aviation Systems CMD, ATTN: AMSAV-GTD, 4300 Goodfellow Blvd, St. Louis, MO 63120-1798
USARTA, Ames Research Center, ATTN: SAVRT-D, Moffett Field, CA 94035-1099
USARTA, Aeroflightdynamics Dir, ATTN: SAVRT-AF, Moffett Field, CA 94035
USARTA, Aviation Applied Technology Dir, ATTN: SAVRT-TY, Ft Eustis, VA 23604
USARTA, Propulsion Dir, ATTN: SAVRT-PN, Lewis Research Center, 21000 Brook Park Rd, Cleveland, OH 44135
USARTA, Aerostructures Dir, ATTN: SAVRT-SD, Hampton, VA 22665
  Avionics R&D Activity, ATTN: SAVAA, Ft Monmouth, NJ 07703-5000
USA Communications-Electronics Cmd, ATTN: AMSEL-TDD, Ft Monmouth, NJ 07703-5001
  Center for Command, Control & Communications System, ATTN: AMSEL-RD-C3, Ft Monmouth, NJ 07703-5202
  Center for Night Vision & Electro-Optics, ATTN: AMSEL-RD-NV-D, Ft Belvoir, VA 22060-5677
  Electronics Warfare/RSTA Center, ATTN: AMSEL-RD-EW-D, Ft Monmouth, NJ 07703-5303
  Signals Warfare Lab, ATTN: AMSEL-RD-SW-D, Vint Hill Farms Station, Warrenton, VA 22186-5100

USA Laboratory Command, ATTN: AMSLC-TD, 2800 Powder Mill Road, Adelphi, MD
   29783-1145
   Atmospheric Sciences Lab, ATTN: SLCAS-DD, WSMR, NM 88002-5501
   Ballistic Research Lab, ATTN: SLCBR-OD, APG, MD 21005
   Electronics Technology & Devices Lab, ATTN: SLCET-D, Ft Monmouth, NJ
      07703-5000
   Electronic Warfare Vulnerability Assessment Lab, ATTN: SLCEW-M-D, WSMR, NM
      8802-5513
   Harry Diamond Labs, ATTN: SLCHD-D, Adelphi, MD 20783-1197
   Human Engineering Lab, ATTN: SLCHE-H, APG, MD 21005-5001
   Materials Technology Lab, ATTN: SLCMT-D, Watertown, MA 02172-0001
   Army Research Office, ATTN: SLCRO-D, RTP, NC 27709-2211
   Army Research, Development & Standardization Group (Europe) Box 65, FPO
      New York 09510
USA Missile Command, ATTN: AMSMI-R, Redstone Arsenal, AL 35898-5240
USA Tank-Automotive Cmd, ATTN: AMSTA-CG, Warren, MI 48397-5000
   Aberdeen Proving Ground, ATTN: STEAP-OC, APG, MD 21005-5001
   Dugway Proving Ground, ATTN: STEDP-CO, Dugway, UT 84022-5000
   Electronic Proving Ground, ATTN: STEEP-DO, Ft Huachuca, AZ 85613-7110
   Aviation Development Test Activity, ATTN: STEBG-CO, Ft Rucker, AL 36362-5276
   Combat Systems Test Act, ATTN: STECS-CO, APG, MD 21005-5059
   White Sands Missile Range, ATTN: STEWS-CG, WSMR, NM 88002-5000
USA Aviation Systems Cmd, ATTN: AMSAV-G, 4300 Goodfellow Blvd, St Louis, MO
   63120-1798
   Belvoir RD&E Center, ATTN: STRBE-Z, Ft. Belvoir, VA 21005-5606
   Natick RD&E Center, ATTN: STRNC-Z, Natick, MA 01760-5000
US Army Corps of Engineers
   Cold Regions Rsch & Eng Lab, ATTN: CRREL-TD, P. O. Box 282, Hanover, NH
      03755-1290
   Construction Eng Rsch Lab, ATTN: CERL-Z, P. O. Box 4005, Champaign, IL
      61820-135
   Engineer Topographic Labs, ATTN: ETL-TD, Ft Belvoir, VA 22060-5546
   Waterways Experiment Station, ATTN: WESTV, P. O. Box 631, Vicksburg, MS
      39180-0631
USA Medical R&D Command, ATTN: SGRD-ZA, Ft Detrick, Frederick, MD 21701-5012
   Aeromedical Rsch Lab, ATTN: SGRD-UAC, Ft Rucker, AL 36362-5292
   Inst of Dental Rsch, WRAMC, ATTN:SGRD-UDZ, Wash, DC 20307-5300
   Inst of Surgical Rsch, ATTN: SGRD-USZ, Ft Sam Houston, TX 78234-6200
   Letterman Army Inst of Rsch, ATTN: SGRD-ULZ, Presidio of San Francisco, CA
      94129-6800
   Biomedical R&D Lab, ATTN: SGRD-UBZ, Frederick, MD 20701-5010
Medical Rsch Inst of Chemical Defense, ATTN: SGRD-UV-ZA, APG, MD 21010-5425
Medical Rsch Inst of Environmental Medicine, ATTN: SGRD-UEZ, Natick, MA
   01760-5007
Medical Rsch Inst of Infectious Diseases, ATTN: SGRD-UIZ-A, Ft Detrick,
   Frederick, MD 21701-5011
Walter Reed Army Inst of Rsch, ATTN: SGRD-UWZ, Washington, DC 20307-5100
USA Health Services Command, Ft. Sam Houston, TX 78234-61000
USA Environmental Hygiene Agency, ATTN: HSHB-Z, APG, MD 21010-5422

USA Research Institute for the Behavioral and Social Sciences, ATTN: PERI-ZT, 5001 Eisenhower Ave, Alexandria, VA 22333-5600
  ARI Field Unit, ATTN: PERI-IJ, POB 2086, Ft Benning, GA 31905
  ARI Field Unit, ATTN: PERI-SB, POB 6057, Ft Bliss, TX 79906
  ARI Field Unit, ATTN: PERI-SH, Ft Hood, TX 76544-5065
  ARI Field Unit, ATTN: PERI-IK, Steele Hall, Ft Knox, KY 40121
  ARI Field Unit, ATTN: PERI-SL, POB 290, Ft Leavenworth, KS 66027
  ARI Field Unit, ATTN: PERI-IR, Ft Rucker, AL 36362-5000
  ARI Field Unit, ATTN: PERI-IO, POB 5787, Presidio of Monterey, CA 93940
USA Training and Doctrine Command, ATTN: ATDO-ZT, Ft Monroe, VA 23651-5000
  Combined Army Test Act, ATTN: ATCT-CG, Ft Hood, TX 76544-5065
COMMANDANTS:
USA Air Defense Cen, ATTN: ATZC-CG, Ft Bliss, TX 79916-5000
USA Armor Cen, ATTN: ATZK-CG, Ft Knox, KY 40121-5000
USA Aviation Cen, ATTN: ATZQ-CG, Ft Rucker, AL 36362-5000
USA Chemical Cen, ATTN: ATZN-CG, Ft McClellan, AL 36205-5000
USA Combined Arms Cen, ATTN: ATZL-SC, Ft Leavenworth, KS 66027-5000
USA Engineer Cen, ATTN: ATZA-CG, Ft Belvoir, VA 22060-5000
USA Infantry Cen, ATTN: ATZB-CG, Ft Benning, GA 31905-5000
USA Intelligence Cen, ATTN: ATSI-SA, Ft Hauchuca, AZ 85613-7000
USA Ordnance Cen, ATTN: ATSL-CMT, APG, MD 21005-5201
USA Signal Cen, ATTN: ATZH-CH, Ft Gordon, GA 30905-5000
USA Field Artillery Center, ATTN: ATZR-C, Ft Sill, OK 73503
SUPERINTENDENT:
US Military Academy, ATTN: Technical Library, West Point, NY 10996
COPIES FURNISHED:
Defense Advanced Research Projects Agency, 1400 Wilson Blvd, Arlington, VA 22209-2308
Defense Logistics Agency, Cameron Station, Alexandria, VA 22304-6183
Defense Technical Information Center, Cameron Station, Alexandria, VA 22304-6145
Lawrence Livermore National Lab, ATTN: L-191, POB 808, Livermore, CA 94550
Los Alamos National Lab, ATTN: Dir for Energy, Rsch & Tech, Los Alamos, NM 87545
NASA HQS, Suite 4237, 400 Maryland Ave, SW, Wash, DC 20546
National Science Foundation, 1800 G Street, NW, Wash, DC 20550
United Nations Library, ATTN: Acquisition Section, Room L-138A, New York, NY 10017
US NAVY
  Naval Air Systems Command, Code AIR-03-D, 1411 Jefferson Davis Hwy, Arlington, VA 22202-3000
  Naval Rsch Lab, ATTN: Dir of Rsch, Washington, DC 20375
  Office of Naval Rsch, Code 10, 800 North Quincy St, Arlington, VA 22217-5000
  HQ US Marine Corps, Code RD-1, Washington, DC 29380-0001
US AIR FORCE
  Air Force Systems Command, ATTN: Technical Director, Andrews AFB, Washington, DC 20334-5000
  Air Force Office of Scientific Research, ATTN: Tech Director, Bolling AFB, Washington, DC 20332-6448

PROCEEDINGS

OF THE

1990 ARMY SCIENCE CONFERENCE

DURHAM, NORTH CAROLINA

12-15 JUNE 1990

VOLUME II

Principal Authors G through M

TABLE OF CONTENTS

PROCEEDINGS OF THE 1990 ARMY SCIENCE CONFERENCE

AUTHOR	TITLE	VOL	PAGE
Crawford, Kenneth	Program Support Environment for the Theater Construction Management System	I	389
Crawford, Robert M.	See Green, Shawn J.	II	81
Cross, Gerald	See Keniston, Richard C.	II	281
Cryz, Stanley	See Sadoff, Jerald C.	III	153
Cymerman, Allen	See Hoyt, Reed W.	II	193
Dagdigian, Paul J.	See Forch, Brad E.	I	573
Dagdigian, Paul J.	See Sausa, Rosario C.	III	225
Dahlstrom, Robert K.	Experimental Development of a Microwave Vlasov Mode Convertor	I	397
Dalrymple, Joel	See Schmaljohn, Connie S.	III	239
Danberg, James E.	Predicted Flight Performance of Base Bleed Projectiles	I	411
Dandekar, Dattatraya P.	Elastic Constants of a Graphite-Epoxy Composite by Ultrasonic Wave Velocity Measurements	I	427
Darrigrand, Andre A.	Reduction of Sweat Accumulation Rate with Pedal Antiperspirants	I	441
Daunt, Geraldine	Investigation of MIR Generation from 1 Micron Lasers Using Non-linear Techniques	I	451
Deberry, J.	See Garvin, Charles	II	27
DeCusatis, C.	See Kwan, H. C.	II	347
DeFatta, Richard P.	See Mullins, William D.	IV	65
DeFrank, Joseph J.	Hydrolysis of Toxic Organophosphorus Compounds by Enzymes from Halophilic Bacteria	I	461
Delgado, Ismael	See Keniston, Richard C.	II	281
DeLuca, Jane P.	See Armstrong, Lawrence E.	I	59
Dempsey, John K.	See Mullins, William D.	IV	65
Ditillo, John T.	See Combs, Roger J.	I	363

AUTHOR	TITLE	VOL	PAGE
Doherty, Tammy J.	See Bruttig, Stephen P.	I	187
Dominessy, Mary E.	See Lukas, Jeffrey H.	II	423
Dornath-Mohr, Michelle A.	Shallow, Low Temperature Au-Ge Ohmic Contacts to GaAs	I	475
Dubois, Doria R.	See Bancroft, William H.	I	73
Dulaney, Marland D., Jr.	Identification of a New Cyanide Countermeasure	I	491
Duncan, Fred	See Keniston, Richard C.	II	281
Dunne, C. P.	See Porter, William L.	III	93
Dutta, Mitra	Semiconductor Optical Waveguide Devices	I	507
Dutta, Mitra	See Choi, K. K.	I	311
Dutta, Mitra	See Cole, Melanie W.	I	349
Dutta, Mitra	See Smith, Doran D.	III	341
Eckart, Donald W.	See Dornath-Mohr, Michelle A.	I	475
Eckels, Kenneth	See Bancroft, William H.	I	73
Enriquez, John I., Sr.	See Keniston, Richard C.	II	281
Evans, Christopher	Metabolic Rescue with a Precursor for Adenylate Synthesis Stimulates Phosphate Utilization and Prolongs Survial in Hemorrhagic Shock	I	517
Falls, Terril C.	Automated Analysis of Interactions Between Terrain and Ground Force Mobility	I	533
Farrand, Timothy G.	See Magness, Lee S.	II	465
Fazi, Christian	Breakdown in Silicon p-n Junctions Due to Short Pulse, High-Field Electromagnetic Stress	I	543
Fedele, Paul D.	Aerosol Transport Modeling Indicates Hazards and Improved Protective Measures for Dusty Agents	IV	1
Ferry, Michael	See Daunt, Geraldine	I	451
Fleetwood, R.	See Bushell, M.	I	227

AUTHOR	TITLE	VOL	PAGE
Swinson, Mark L.	Military Robotics: A Technology Assessment From a User's Perspective	III	423
Szafraniec, Linda L.	See Yang, Yu-Chu	III	547
Szlyk, Patricia C.	An Innovative Fluid Delivery System for Chemical Protective Clothing: An Evaluation	III	437
Tanton, George A.	UV-IR Detector and Focal Plan Array Material Evaluation Using Faraday Rotation	III	447
Taub, Irvin A.	See Kim, Hie-Joon	II	295
Thompson, Donald L.	See Chabalowski, Cary F.	I	295
Thompson, Thomas J.	See Pleban, Robert J.	III	69
Trevino, S. F.	A Determination of the Inter-molecular Potential of Nitromethane	III	461
Underwood, Robert B.	See Falls, Terril C.	I	533
Valdes, James J.	Direction of Toxins with a Reversible Biosensor	III	477
Valentine, Patrick J.	See Pleban, Robert J.	III	69
Vanderhoff, John A.	Simultaneous Determination of Temperatures of OH Concentrations in a solid Propellant Flame	III	491
Vlahocos, Constaine	See Dulaney, Marland D., Jr.	I	491
Vodkin, M. H.	See Hoover, Timothy A.	II	181
Vogel, James A.	See Friedl, Karl E.	I	631
Wade, Charles E.	See Bruttig, Stephen P.	I	187
Walker, Clinton	See Busciglio, Henry H.	I	213
Walker, John E.	See Halliday, John W.	II	135
Walker, John E.	See Steeves, Diane M.	III	395
Walkinshaw, John W.	See Perkins, Janet	III	55

AUTHOR	TITLE	VOL	PAGE
Ward, Alford L.	See Fazi, Christian	I	543
Weinacht, Paul	Navier-Stokes Predictions of Pitch Damping for High L/D Finned Kinetic Energy Projectiles	III	503
Weinacht, Paul	See Guidos, Bernard J.	II	119
Weiner, M.	See Kim, Anderson	II	307
Weiner, Maurice	See Braun, Christopher G.	I	145
Westgate, Roger C.	See Fazi, Christian	I	543
Wheeler, Thomas J.	See Richardson, Judith D.	III	109
White, Leonard A.	Setting Enlistment Standards on the ABLE to Reduce Attrition	III	519
Wiesmann, William	See Evans, Christopher P.	I	517
Williams, Jim C.	See Hoover, Timothy A.	II	181
Willingham, Reginald A.	See Singler, R. E.	III	315
Wisher, Robert A.	Predicting the Decay of Mobile Subscriber Equipment (MSE) Operator Skills	III	533
Wisler, John A.	See Dulaney, Marland D., Jr.	I	491
Wisniewski, Henry L.	See Santiago, Joseph M.	III	179
Wolfe, A. David	See Smejkal, Ruthann M.	III	329
Wood, Gary L.	See Sharp, Edward J.	III	263
Woontner, Susan	See O'Neill, Timothy R.	III	15
Wren, Gloria P.	See Coffee, Terence P.	I	323
Yang, Yu-Chu	Oxidative Decontamination of the Chemical Nerve Agent	III	547
Yip, Pearl W.	See Perkins, Janet	III	55
Youmans, R.	See Kim, Anderson	II	307
Younger, M.	See Garvin, Charles	II	27
Zeto, R.	See Kim, Anderson	II	307

The Validation and Application of a Rotor Acoustic Prediction Computer Program

Judith Hoeffner Gallman
U.S. Army Aeroflightdynamics Directorate
NASA Ames Research Center, Mail Stop 215-1
Moffett Field, California 94035

Introduction

The rapid development of passive acoustic arrays and signal processing technology that detects, tracks, and identifies a rotorcraft poses a growing threat to the survivability of U. S. combat helicopters. An essential prerequisite to reducing the acoustic detectability of military rotorcraft, and hence to increase their survivability, is a better understanding of main rotor noise, which is the major contributor to the overall noise. A simple, yet accurate, Rotor Acoustic Prediction Program (RAPP) has been developed to advance the understanding of main rotor noise. This prediction program utilizes the Ffowcs Williams and Hawkings (FW-H) equation. The particular form of the FW-H equation[1] used in this analysis is well suited for the coupling of the measured blade surface pressure to the prediction of acoustic pressure.

The FW-H equation is an inhomogeneous wave equation that is valid in all space and governs acoustic pressure generated by thin moving bodies. The nonhomogeneous terms describe mass displacement due to surface motion and forces due to local surface stresses, such as viscous stress and pressure distribution on the surface. There is a rich history of the application of the FW-H equation as applied to the prediction of propeller and rotor noise. The events of the past, which have contributed to the wealth of knowledge in which this paper has its foundation, are discussed in depth in references 1 and 2. The history begins with Gutin's[3] compact source assumption. He modeled steady aerodynamic forces on a propeller with harmonic acoustic dipoles in the frequency domain. Garrick and Watson[4] extended this approach to uniform rectilinear motion. Deming,[5] Arnoldi,[6] and Lyon[7] studied the effects of blade thickness on acoustic noise respectively for static propellers, uniform rectilinear motion, and for helicopters in forward flight. Farassat,[1] Hawkings and Lowson,[8] Isom,[9] and Schmitz and Yu[2] have used noncompact monopole terms to model rotor blade thickness and distributed dipoles to model local blade surface forces.

There are four types of main rotor noise: BVI noise, low-frequency noise, high-speed impulsive noise, and broadband noise.[10] Blade-vortex interaction noise occurs when a tip vortex, previously shed by a rotor blade, passes close enough to a rotor blade to cause large

variations in the blade surface pressures. This event is most disturbing when it happens on the advancing side of the rotor disk.[11] Low-frequency noise includes hover and low to moderate speed forward flight. For these flight conditions, the low frequency components of the acoustic signal dominate. RAPP was developed for BVI and low-frequency applications. These are the flight regimes producing the propagating noise that leads to acoustic detection.

High-speed impulsive noise is a highly nonlinear, transonic phenomenon that is beyond the scope of this work. Recent efforts that do address high-speed impulsive noise couple computational fluid dynamics (CFD) with acoustics. The particular method of Isom and Purcell[12] couples nearfield pressure from a full potential finite-difference method with a Kirchhoff integral formulation to extend the CFD results to the far field. Broadband noise will not be addressed in this paper either. It is of comparatively small amplitude and is a result of unpredictable blade loading such as that caused by turbulence.[13]

RAPP implements the noncompact methods to efficiently and accurately predict BVI noise and low-frequency noise characteristics generated by the main rotor. This paper will explain the details of the blade thickness modeling and the blade surface force modeling used in RAPP. It will establish the validity of RAPP by comparisons with aeroacoustic experimental data for BVI and low-frequency noise. Future applications that are made possible by prediction programs such as RAPP are discussed. The most important application is to use RAPP in conjunction with an aerodynamic blade surface pressure prediction program to learn to design quieter rotor blades. Hence, this would minimize detection ranges and decrease the threat to the survivability of combat helicopters.

Development of the Rotor Acoustic Prediction Program

As stated earlier, RAPP is based on the FW-H equation. This integral equation is expressed as follows:

$$
4\pi p'(\vec{x}, t) = \frac{\partial}{\partial t} \int \left[\frac{\rho_0 v_n}{r|1 - M_r|} \right]_{ret} dS(\vec{y})
$$

$$
+ \frac{1}{a_0} \frac{\partial}{\partial t} \int \left[\frac{p_i \hat{r}_i}{r|1 - M_r|} \right]_{ret} dS(\vec{y}) + \int \left[\frac{p_i \hat{r}_i}{r^2|1 - M_r|} \right]_{ret} dS(\vec{y}) \qquad (1)
$$

Equation (1) is an integral equation for the acoustic pressure, p', at a point $\vec{x}$ and a time t. The subscript ret indicates that the integrals are to be evaluated at the retarded time $\tau = t - r/a_0$. The speed of propagation of sound in the fluid medium is designated by a_0, and ρ_0 is the density of the undisturbed fluid medium. M_r is the Mach number in the radiation direction, $M_r = \frac{V_i \hat{r}_i}{a_0}$. In this equation for the radiation Mach number, V_i is the local free stream velocity vector, and $\hat{r}_i$ is the unit vector describing the radiation direction. For the above form of equation (1), the radiation Mach number should not equal 1. As a rule, the advancing tip Mach number should remain below the critical Mach number for the airfoil section. RAPP does not contain any nonlinear terms and should not be applied to nonlinear circumstances such as the transonic flow regimes.

The first term on the right-hand side of equation (1) is referred to as the thickness term. This term represents the disturbance of the fluid medium caused by the airfoil. For a fixed volume, a short fat airfoil will cause a greater displacement than a long slender airfoil. This is the influence of the normal velocity on the airfoil surface. Thin airfoil theory defines the normal velocity as

$$v_n = V_0 [\alpha + \tan^{-1} \frac{dy}{dx}]$$ (2)

Here α is the airfoil angle of attack, V_0 is the local freestream velocity, and $\frac{dy}{dx}$ is the thickness distribution of the airfoil. Even though the angle of attack is not zero, the term

$$\frac{\partial}{\partial t} \int \frac{\rho_0 V_0 \alpha}{r(1 - M_r)} dS(\vec{y})$$ (3)

is negligible. It does not contribute to the thickness noise, even for $\alpha \geq 10°$, and has been neglected by other researchers, namely Schmitz and Yu[2]. The term

$$\frac{\partial}{\partial t} \int \frac{\rho_0 V_0 \frac{dy}{dx}}{r(1 - M_r)} dS(\vec{y})$$ (4)

can be approximated quite well by a chordwise distribution of sources and sinks at about five spanwise locations. The sources are assumed to be of strength

$$ss = \rho_0 V_0 \frac{dy}{dx}$$ (5)

Equation (4) can be discretized as

$$\rho_0 \frac{\partial}{\partial t} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{V_{0_{ij}} dy_{ij}}{r_{ij}(1 - M_{r_{ij}})} dz_i$$ (6)

Here $dS = dxdz$, and the subscript ij indicates quantities that are functions of both the chordwise and spanwise variables. The thickness of the airfoil at the ith spanwise location and the jth chordwise location is dy_{ij}. The thickness noise has been well modelled with 10 chordwise sources at each of 5 spanwise radial stations. Adding more sources and more stations does not improve the accuracy of the thickness prediction. It just takes more computational time. This was verified by adding more radial stations and more sources at these radial stations to RAPP in the developmental stage.

The second and third terms in equation (1) are the loading terms. Only the second term in equation (1) will be used to predict the loading noise for two reasons. The third term, which does not include a partial time derivative, is considered the near-field term. It is divided by the square of the distance between the source and the observer. For an observer in the far-field, this near-field term is negligible. Also, the quasi-steady near-field

acoustic pressure was not measured by the microphones used in acquiring acoustic test data. These microphones were back vented free-field response microphones. Consequently the frequency of the quasi-steady near-field acoustic pressure was below the low frequency response of these microphones.

RAPP uses the acoustic lifting line formulation to model the loading noise. If the loading information is supplied as blade surface pressure, the sectional lift and drag are determined by integrating the surface pressure over the chord. The loading term can be discretized as follows:

$$\frac{1}{a_0}\frac{\partial}{\partial t}\sum_{i=1}^{m} c_i \left[\frac{l\hat{n}_l \cdot \hat{r} + d\hat{n}_d \cdot \hat{r}}{r(1 - M_r)}\right]_i dz_i \tag{7}$$

The sectional lift and drag are defined by l and d. The normal vectors $\hat{n}_l$ and $\hat{n}_d$ are defined parallel and perpendicular to the tip path plane of the rotor. The airfoil is exerting a force on the fluid, therefore, convention defines lift as positive downward and drag as positive toward the leading edge of the airfoil. Figure (1) displays the sign convention of these normal vectors in the blade fixed coordinate system. Two important assumptions that have been made in arriving at equation (7) are that the distance r and the radiation Mach number M_r are held constant over the chord, c_i. The loading terms are modelled well by equation (7) with five radial stations on the outer 50 percent of the rotor span.
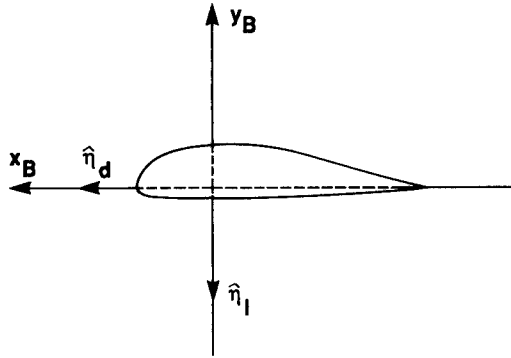
Figure 1. Presentation of the sign convention for the surface normals in the blade fixed coordinate system.

The thickness and loading terms must be evaluated at the retarded time, τ. This allows for the proper summation of the acoustic disturbances at the observer location and time. An iterative technique is needed to solve for the location of all the sources that contribute to the acoustic disturbance for a time t and observer $\vec{x}$. Iteration is necessary to solve the retarded time equation because the distance between the source and the observer is a function of the source time. The locations of the contributing sources comprise the acoustic planform. The acoustic planform is defined as the locus of emitting source locations whose signals arrive simultaneously at a time t to the observer. Please note that the surface over which

the integration is carried out is not the acoustic planform, but rather the physical geometric planform.

Both the Newton iteration method and the Bisection method are used to solve for the source locations on the acoustic planform. Newton's method converges quickly, or in just a few function evaluations, if the initial guess is "close enough" to the solution. Although the function on which the iterations are being performed,

$$f(x) = t - (\tau - \triangle\tau + r(\tau - \triangle\tau)/a_0) \qquad (8)$$

is a continuously increasing function, $f\prime(x)$ can become very small. Since this term is the denominator in the iteration scheme, the method will not converge unless the initial guess is close enough to the solution. To guarantee convergence of Newton's method, a satisfactory initial guess is supplied by the Bisection method. This method is not affected by $f\prime(x)$, but it does require many function evaluations. This is why Newton's method is used to conclude the iteration after the Bisection method finds a suitable first guess.

A central difference method, which is second-order accurate, is used for the time derivative. The time derivatives of the thickness and loading terms are taken separately, so that the contribution to the acoustic disturbance from each phenomenon can be studied separately.

Validation of the Prediction Code

Discussion of Test Data

RAPP was validated by comparisons with the Boeing 360 experimental acoustic test data acquired during the Aerodynamic and Acoustic Testing of Model Rotors (AATMR) program at the Duits-Nederlandse Windtunnel (DNW). The DNW Boeing 360 model rotor acoustic database has been reviewed and analyzed by Zinner et al.,[10] and is considered of high quality to validate acoustic prediction codes. Zinner shows that the Boeing 360 acoustic test data displays the expected trends for high-speed impulsive noise, blade-vortex interaction noise, low-frequency noise, and broadband noise. The conditions for maximum BVI noise were shown to range from low shaft-tilt angles at high advance ratios to high shaft-tilt angles at low advance ratios. BVI noise has maximum radiation in the direction toward the advancing side of the rotor and at least 25 degrees below the rotor plane. A severe test of RAPP's predictive ability will be whether or not it can predict these trends for BVI.

The accuracy of RAPP's predictive ability depends on the accuracy of the input blade surface pressure. The DNW Boeing 360 experimental surface pressure data has been utilized by Caradonna et al[14] in the prediction of airloads and is considered to be of the quality necessary to support test/theory. Thus, this DNW Boeing 360 blade surface pressure was thought to be a good source of input to validate RAPP. Both Caradonna and Zinner have discussed the DNW test facility, the data acquisition system, and the characteristics of the model rotor in great detail. The interested reader should study references 10 and 14 for more information on this experiment.

The sectional lift and drag, supplied as input for RAPP, were obtained by integrating the Boeing 360 surface pressure over the chord. This technique did not work well for the

hover flight regime since drag dominates the loading noise contribution. There are an insufficient number of pressure taps to get an accurate estimate of the sectional drag. Analytical estimates of the sectional drag are better suited for this application and are easy to determine for hover where the loading is ideally constant about the azimuth. A good source of the sectional lift and drag is a potential flow lifting surface analysis program, such as that by Analytical Methods Incorporated (AMI).[15] The sectional lift and drag from AMI's hover code was used in RAPP to predict the helicopter noise in hover. For the forward and descending flight regimes, the DNW Boeing experimental surface pressure data was used as input. The technique of integrating the surface pressure chordwise to get the sectional lift and drag worked well for these flight regimes because the lift dominates the loading noise contribution.

Figure 2 shows the microphone array that was used to acquire test data. These microphone locations define the observer positions. This array of microphones is located a distance of three rotor radii from the rotor hub.
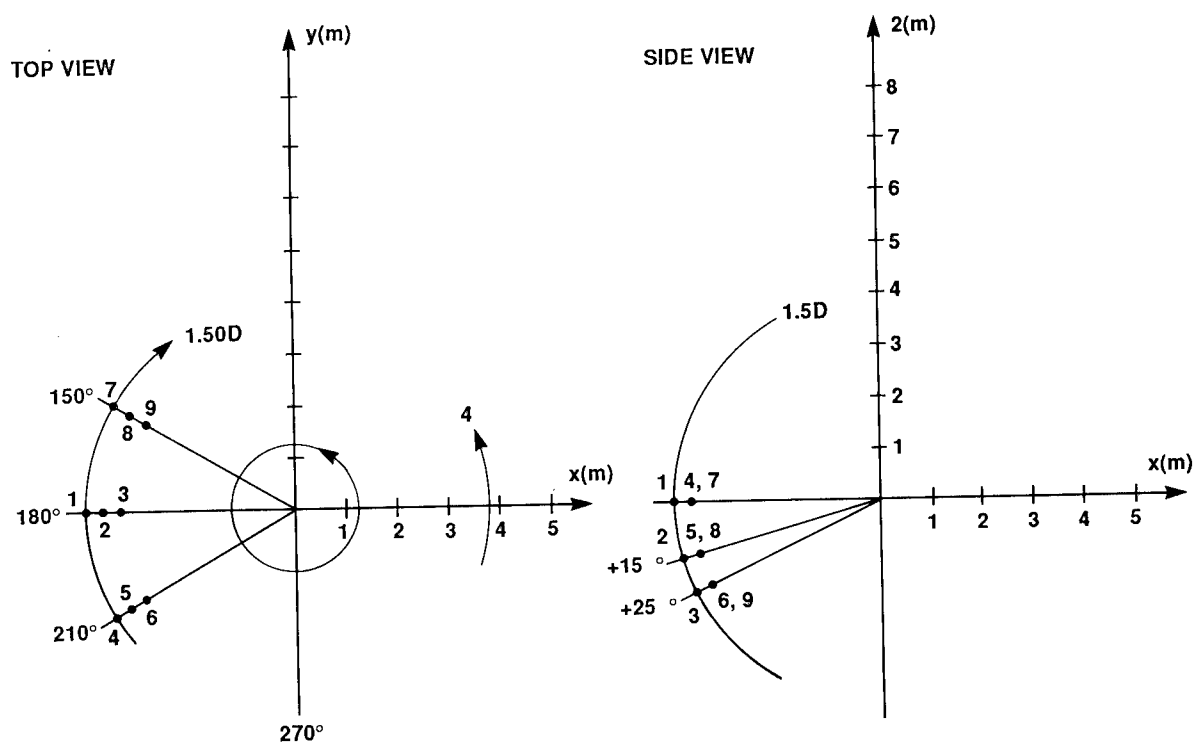
Figure 2. Microphone array used to acquire acoustic pressure for the Boeing model 360 DNW/AATMR test.

Discussion of Comparisons

RAPP was first applied to predict acoustic pressure in the hover flight regime. Since the loading is ideally constant around the azimuth for the hovering rotor, it was a natural first step in the developmental process. Five hover test points were chosen from the Boeing 360 data base for the hover comparisons. All five test points had the same solidity weighted thrust coefficient $C_T/\sigma = .07$, with hover tip Mach number ranging from $M_H = .5$ to $M_H = .636$. The loading, or the sectional lift and drag, from the AMI Hover code supplied input to RAPP to match these five test conditions. In Figure 3a, the plots of predicted peak negative pressure versus Mach number depict the trend for hover of decreasing peak negative pressure below the rotor tip path plane. The peak negative pressure is defined as the most negative peak amplitude in the acoustic pressure time histories. At this loading condition, the thickness noise is dominant and its amplitude decreases below the tip path plane. The loading noise, which becomes predominant below the plane, does not increase more than the thickness decreases. Predicted peak negative pressure versus Mach number plots are compared to the measured plots of peak negative pressure versus Mach number in figures 3b, 3c, and 3d. Figure 3b shows comparisons for the in-plane microphone, $\phi = 0°$. The comparisons for the in-plane microphone are quite acceptable. Figure 3c shows comparisons for the microphone at $\phi = 15°$ below the tip path plane, and figure 3d shows comparisons for the microphone at $\phi = 25°$ below the tip path plane. The comparisons in figures 3c and 3d are not acceptable. The experimental plots do not display the expected trend of decreasing below the tip path plane. This is due to recirculation of the air in the DNW test chamber. The DNW is not a hover chamber and does not allow for clean hover test conditions. A mean flow and recirculation occur in the test chamber and cause unsteady effects in the loading and in the acoustics. These unsteady affects are most noticeable in the out-of-plane observer locations and are quite noticeable in the plots shown in figures 3c and 3d. However, the in-plane comparisons, as shown in figure 3a, are good enough to conclude that RAPP can accurately predict acoustic pressure for the hover flight regime.

The next step in the development of RAPP was to predict acoustic pressure in forward flight and in descent where BVI occurs. In these flight regimes, RAPP will have to contend with blade loading that varies about the azimuth. To get an accurate prediction, the blade loading must be known at many locations around the azimuth. The measured experimental pressure is known at 1024 locations around the azimuth. This resolution of almost every one-third of a degree enables the acoustic prediction program to compute accurate pressure time histories. Computer programs that calculate blade surface pressure for forward flight and BVI flight regimes are beginning to have the ability to predict the blade loading at very fine increments in azimuthal angle. The application of this capability to the Boeing 360 rotor will be discussed in the application section of this paper.

To validate RAPP for the forward and descending flight regimes, the measured surface pressure data from the Boeing 360 model rotor test was used as input. The surface pressure was integrated over the chord to get the sectional lift and drag at each radial station and azimuth angle. The acoustic pressure time histories, predicted using the experimental surface pressures as input to RAPP, compare well with the measured acoustic pressure time histories. This is presented in figures 4 and 5 which show comparisons of predicted to measured acoustic pressure time histories. Figure 4a shows the in-plane comparisons for a moderately high speed forward flight test point. Figure 4b shows comparisons for observers 25° below

7
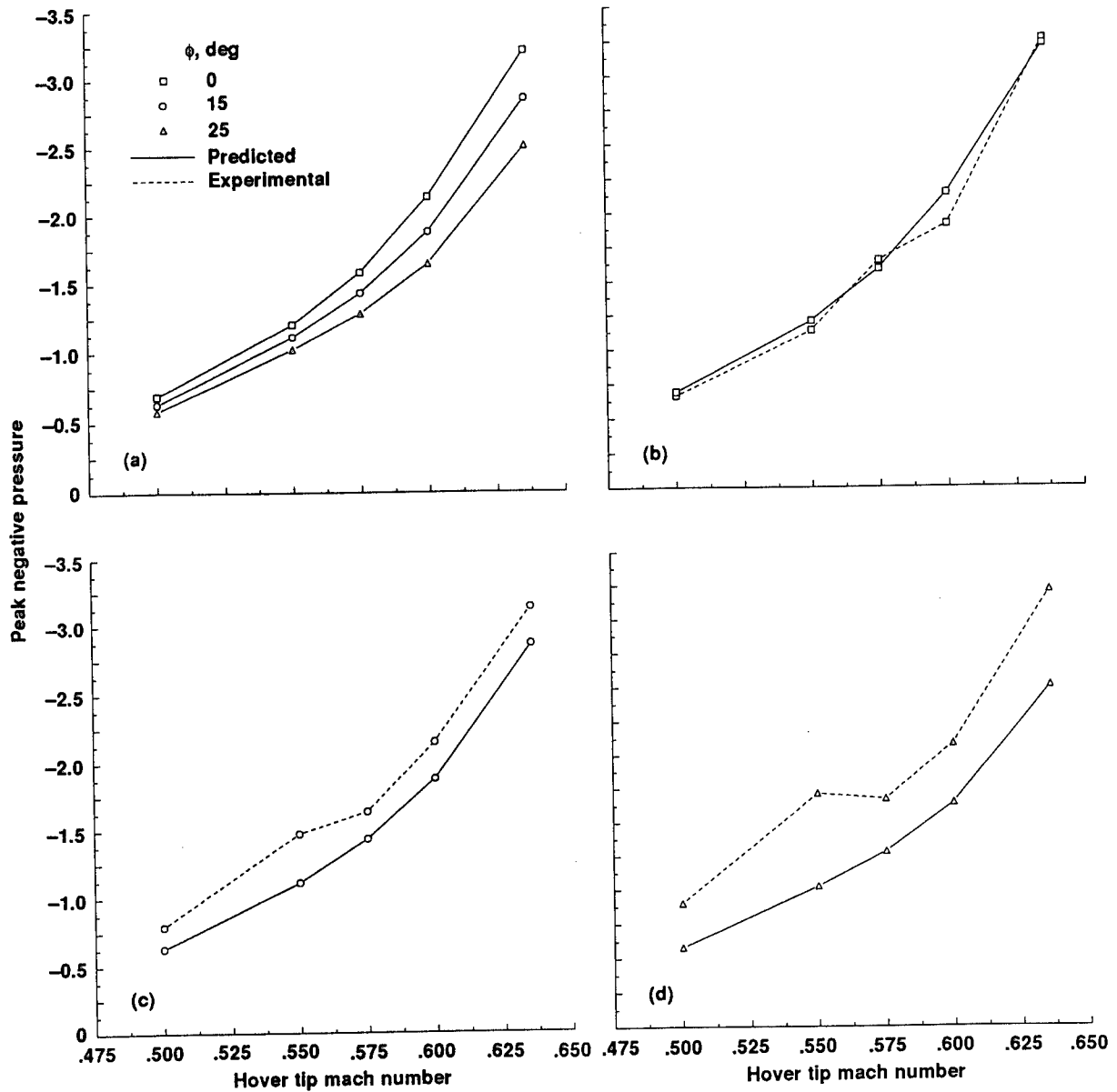
Figure 3. Plots of peak negative acoustic pressure versus mach number. (a) Depicts predicted trends. (b) Shows comparisons for inplane microphones, $\phi = 0$. (c) Shows comparisons for microphones at $\phi = 15°$ below the tip path plane. (d) Shows comparisons for microphones at $\phi = 25°$ below the tip path plane.

the tip path plane for this same test point. Figure 5a shows the in-plane comparisons for a descent or BVI test point. Figure 5b shows comparisons for observers 25° below the rotor plane. This data for BVI portrays the expected trends where the highest peak-to-peak levels

8
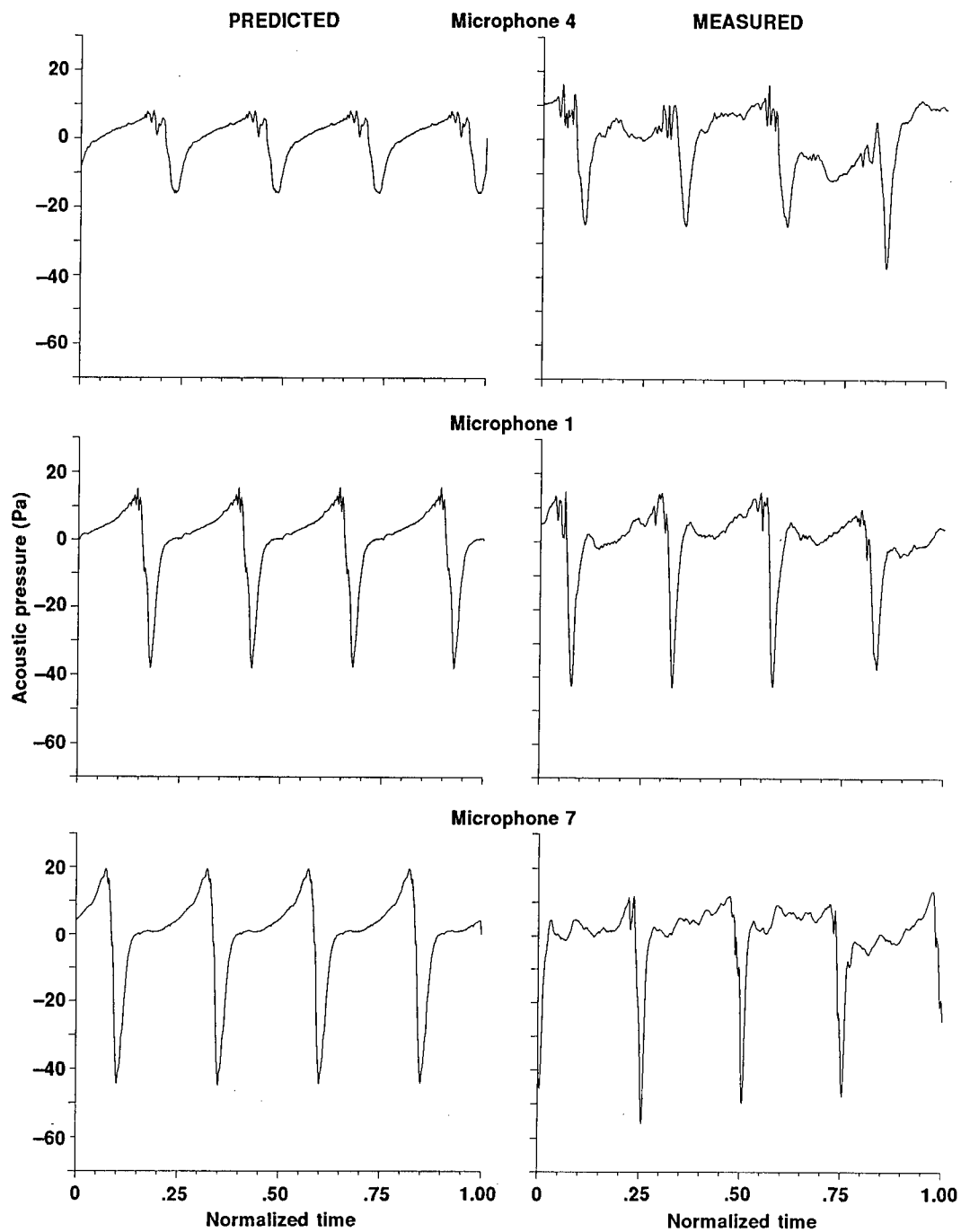
Figure 4a. Plots of acoustic pressure time histories for microphones in the tip path plane. The flight conditions are $\mu = 0.356$, $CT/\alpha = 0.07$, and shaft tilt $=-5.55°$.

9

PREDICTED   Microphone 6   MEASURED

Microphone 3
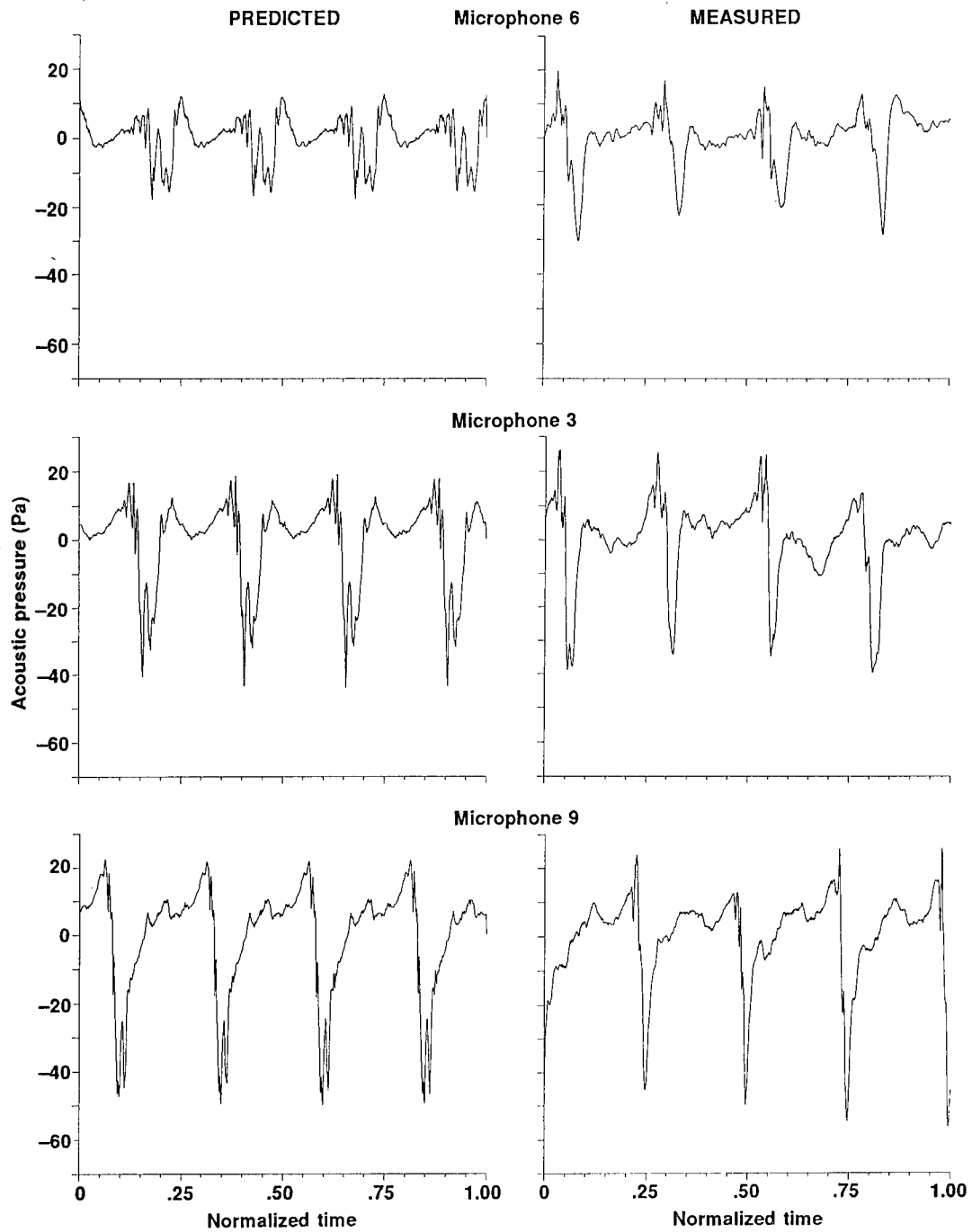
Microphone 9

Figure 4b. Plots of acoustic pressure time histories for microphones 25° below the tip path plane. The flight conditions are $\mu = 0.356$, $CT/\alpha = 0.07$, and shaft tilt $= 5.55°$

10
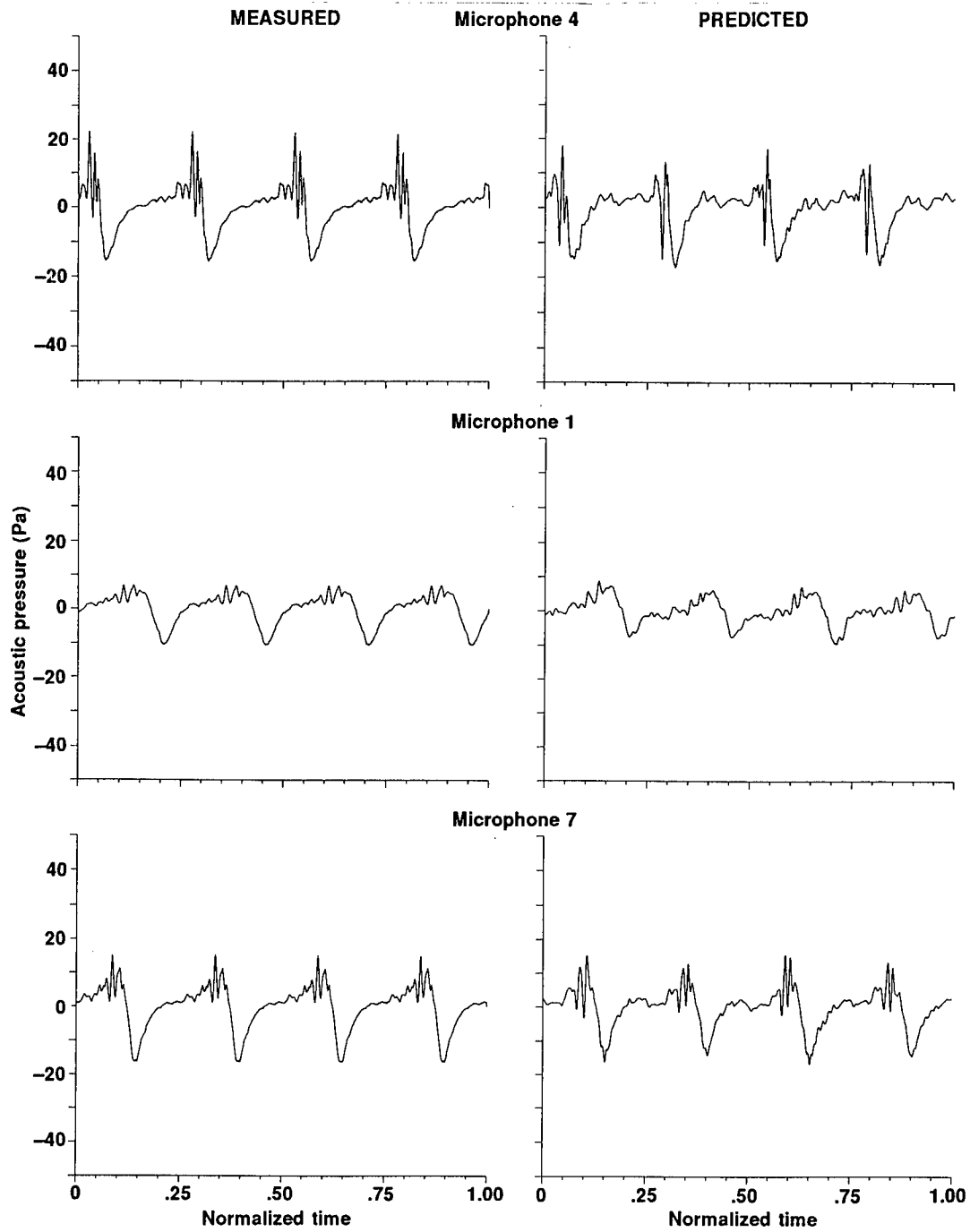
Figure 5a. Plots of acoustic pressure time histories for microphones in the tip path plane. The flight conditions are $\mu = 0.225$, $CT/\alpha = 0.07$, and shaft tilt $= 4.38°$.
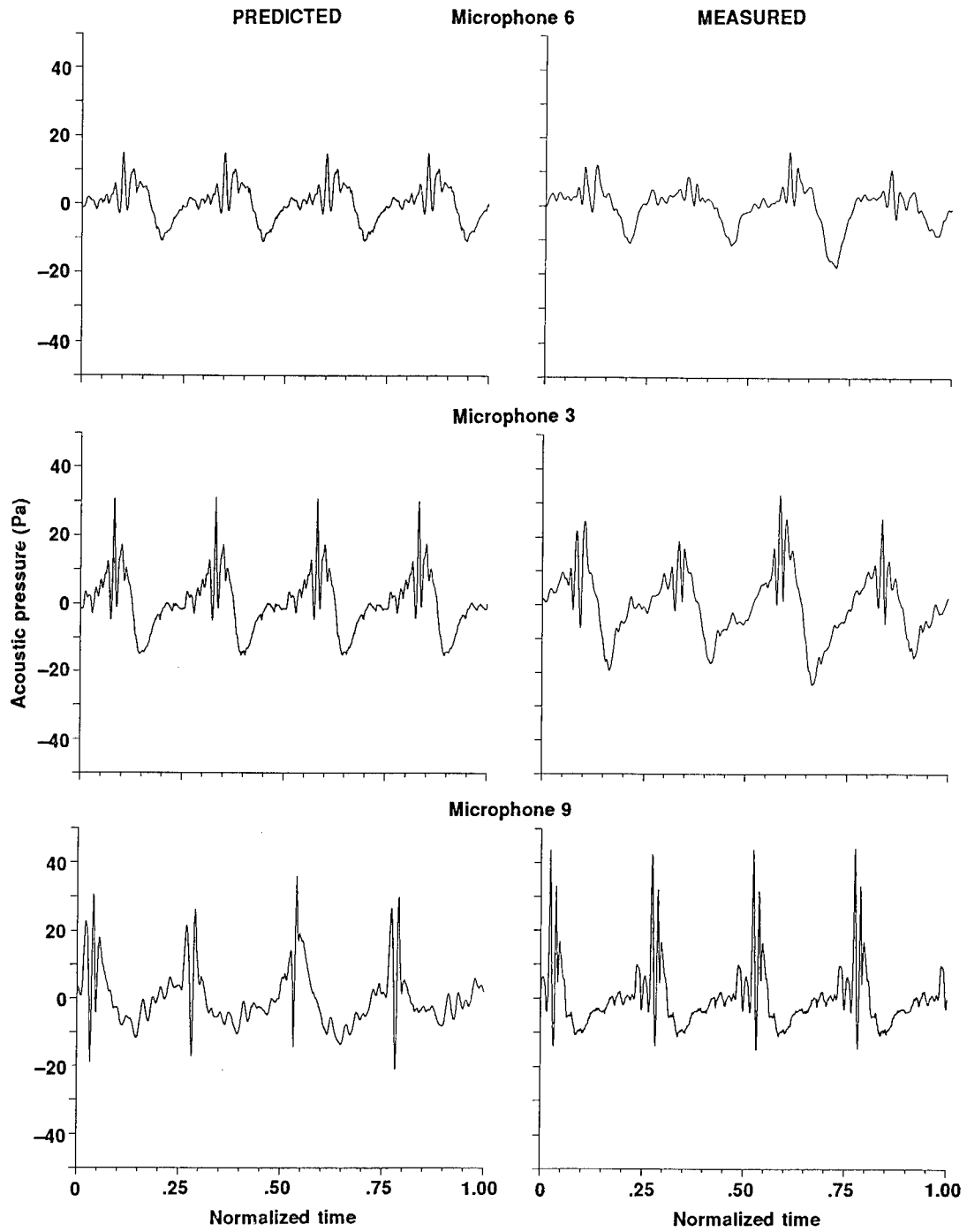
Figure 5b. Plots of acoustic pressure time histories for microphones 25° below the tip path plane. The flight conditions are $\mu = 0.225$, $CT/\alpha = 0.07$, and shaft tilt = 4.38°.

are near 150° azimuth and 25° below the plane. It can be concluded from the comparison plots in figures 4 and 5 that RAPP is capable of predicting accurate acoustic pressure time histories for moderate-speed forward flight and BVI.

Applications

A prediction program such as RAPP, can be applied to study many aspects of main rotor noise. The contribution of the three noise generating mechanisms— thickness, lift, and drag— to the total acoustic pressure disturbance can be plotted separately. This allows for a more in-depth study of each noise generating mechanism. An example of such a plot is shown in figure 6 for an in-plane microphone from the same forward flight test point shown in figure 4a.

The prediction of acoustic pressure is dependent upon the knowledge of the blade loading. Reliable sources of predicted blade loading are becoming available as the research in blade surface pressure prediction progresses. The most promising technique developing is the coupling of a full-potential rotor flow solver with a comprehensive lifting surface analysis. It is becoming possible for the techniques to produce the surface pressure at the number of azimuthal angles necessary for accurate acoustic prediction for forward flight and BVI. A very fine resolution in azimuth angle is necessary for accurate BVI predictions because the surface pressure changes so abruptly in a short source time increment. Possible future applications for RAPP are to use the blade surface prediction capabilities to study the effects of varying airfoil sections, twist distributions and planform shape on BVI. This application could lead to helicopter rotor systems that are designed to be quieter in the BVI flight regime.

Another application of RAPP, when used in conjunction with blade surface pressure prediction codes, is to predict experimental wind tunnel test data before the test is run. This would allow the experimentalist to decide the best microphone specifications and locations and at which test conditions to run. These predictions of the test data could help the experimentalist decide on the integrity of the data being acquired. The overall effect would be better planned wind tunnel tests that would provide more useful data.

Conclusions

The most important conclusion that can be drawn from the information presented in this paper is that RAPP can predict accurate rotor acoustic pressure time histories. The notable technique and assumption used in RAPP is the acoustic lifting line formulation. This formulation works well and considerably simplifies the application of the FW-H equation. RAPP is a simple enough program to run on a personal computer and executes in RAP-Pidly. Therefore, the above mentioned applications could take place in the test field or in an industrial setting where supercomputers are not the norm.

Some interesting byproducts of RAPP's developmental process concern the details of the computer program. Details such as 10 sources along the chord at 5 radial stations being necessary and sufficient to accurately predict thickness noise. It is also interesting that even BVI noise can be predicted accurately by modeling the loading terms at only 5 radial stations. This indicates that the azimuthal resolution is more important than the
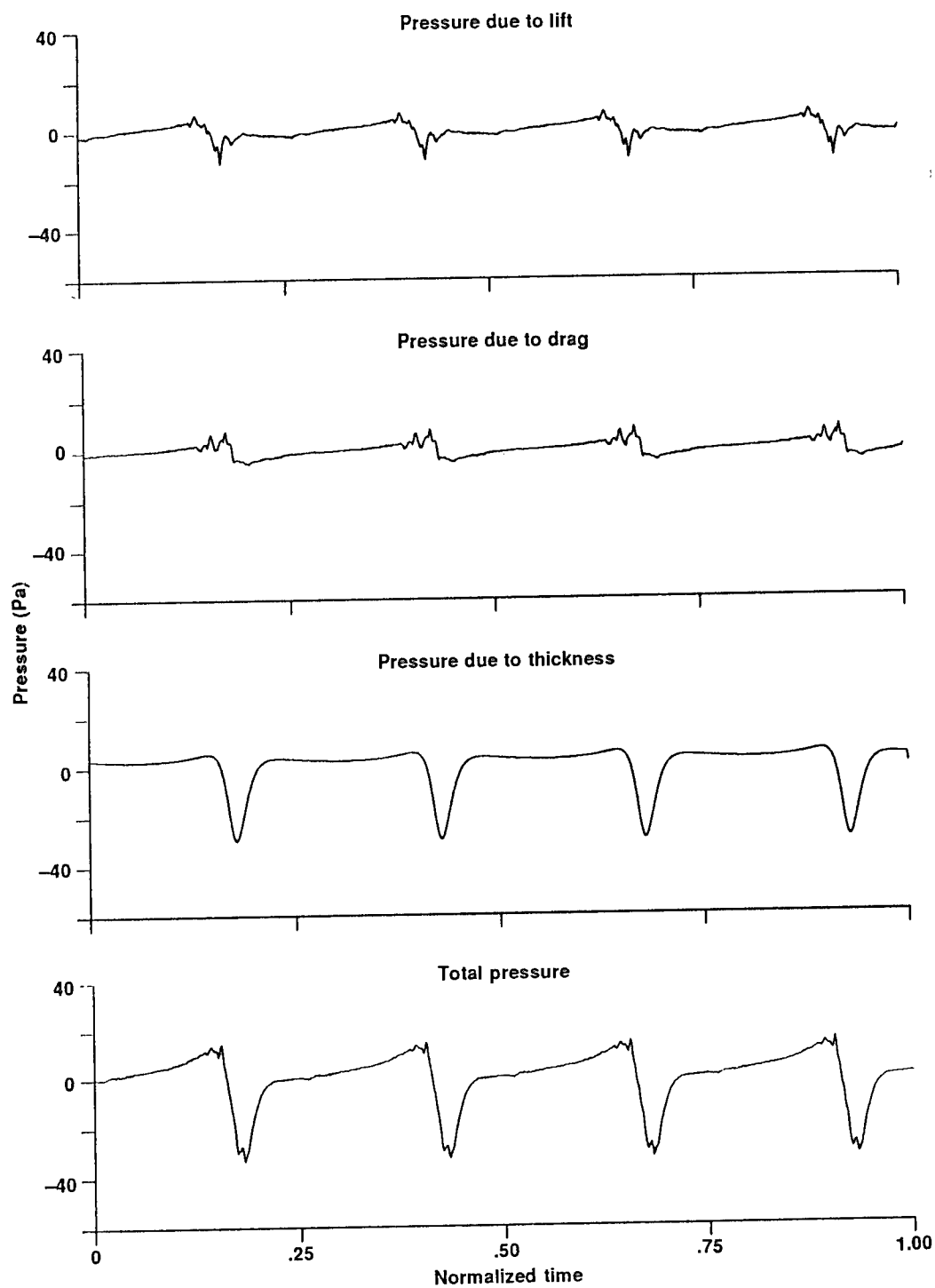
Figure 6. The contributions of the noise generated by 11f, drag and thickness are plotted separately.

spanwise resolution. The fact that the acoustic predictions for hover rely so heavily on accurate section drag input will prove useful in future hover acoustic predictions.

References

[1] F. Farassat, Theory of Noise Generated From Moving Bodies with an Application to Helicopter Rotors, *NASA TR R-452*, Dec. 1965.

[2] F. Schmitz, Y. H. Yu, Theoretical Modeling of High Speed Helicopter Impulsive Noise, Presented at The 3rd European Rotorcraft and Powered Lift Aircraft Forum, Aux-En-Provence, France, Sept. 1977.

[3] L. Gutin, On the Sound Field of a Rotating Propeller, *Physik. Zeitschr. der Sowjetunion*, Bd. 9, Heft 1, 1936, pp. 57-71. Trans. *NACA TM 1195*, 1948.

[4] I. E. Garrick, C. E. Watkins, A Theoretical Study of the Effect of Forward Speed on the Free-Space Sound Pressure Field Around Propellers, *NACA Report 1198*, 1954.

[5] A. F. Deming, Noise from Propellers with Symmetrical Sections at Zero Angle, *NACA TN 679*, 1938.

[6] R. A. Arnoldi, Propeller Noise Caused by Blade Thickness, Rep. R-08961, Res. Dep., United Aircraft Corp., Jan. 10, 1956.

[7] R. H. Lyon, Radiation of Sound by Airfoils that Accelerate Near the Speed of Sound, *J. Acoustical Society of America*, Vol. 49, No. 3, March 1971, pp. 894-905.

[8] D. L. Hawkings, M. V. Lowson, Theory of Open Supersonic Rotor Noise, *J. Sound and Vibration*, Vol. 36, No. 1, Sept. 1974, pp. 1-20.

[9] M. P. Isom, The Theory of Sound Radiated by a Hovering Transonic Helicopter Blade, Poly-AE/Am. Rep. No.75-4. Department of Aerospace Engineering and Applied Mechanics, Polytechnic Institute of New York, May 1975.

[10] R. A. Zinner, et al., Review and Analysis of the DNW/Model 360 Rotor Acoustic Data Base, Presented at The 15th European Rotorcraft Forum, Amsterdam, Netherlands, Sept. 1989.

[11] F. H. Schmitz, Y. H. Yu, Helicopter Impulsive Noise: Theoretical and Experimental Status, *J. Sound and Vibration*, Vol. 109, 1986, pp. 361-422.

[12] M. Isom, T. Purcell, R. Strawn, Geometrical Acoustics and Transonic Helicopter Sound, Presented at the AIAA 11th Aeroacoustics Conference, Sunnyvale, California, Oct. 1987.

[13] T. F. Brooks, et al., Main Rotor Broadband Noise Study in the DNW, AHS Aerodynamics and Aeroacoustics Specialists Meeting, 1987.

[14] L. Dadone, F. Caradonna, et al., The Prediction of Loads on the Boeing Helicopter's Model 360 Rotor, Presented at The 45th Annual Forum and Technical Display of the American Helicopter Society, Boston, May 1989.

[15] J. M. Summa, B. Maskew, A Surface Singularity Method for Rotors in Hover or Climb, *USAAVRADCOM TR 81-D-23*, December 1981.

Induction of Enhanced Strains of a Thermophile That Synthesize
an OPA Anhydrase Effective in Hydrolyzing 4-Nitrophenyl
Esters of Phenylphosphinate

Benedict J. Gallo, Dr., Paula M. Scotland, Ms., and David
A. Gowenlock, Mr.
U.S. Army Research, Development and Engineering Center,
Natick, MA 01760-5020

Introduction

The bacterial thermophile JD100, a soil microbe, was identified as a strain of Bacillus stearothermophilus and has been reported to synthesize a unique endogenous organophosphorus acid anhydrase (OPA anhydrase), which cannot hydrolyze diisopropylfluorophosphate (DFP) but is very active in hydrolyzing 4-nitrophenyl esters of phenylphosphinate.[1] The use of this and other similar OPA anhydrases in hydrolyzing organophosphorus esters and pesticides has been suggested by a number of authors[2]. In order to make enzyme from this strain of B. stearothermophilus more available for research and development, we conducted a program to produce hyperproducing mutants. This paper describes the methodology and results from our studies.

Experimental methods

Organisms - Bacillus stearothermophilus JD100, JD200 and JD300 were obtained from Dr. J. DeFrank, U. S. Army Chemical Research, Development and Engineering Center (CRDEC). They were maintained on LB (Lennox) agar slants at 55°C for 24-36 hours before storage at refrigerator temperatures. The streptomycin resistant (strE) strains of Bacillus stearothermophilus reported in this paper were put on the above medium but with the addition of 250 ug to 500 ug of streptomycin per mL medium.

Mutation and Mutant Screening - Log phase cells of Bacillus stearothermophilus grown at 45°C were mutagenized with UV light at room temperatures at a kill rate in excess of 99.99%. Immediately after irradiation survivors were protected from light, cooled and stored at 4°C for 18 to 36 hours prior to use. The mutagenized cells were put on agar plates to a average colony density of 100 per plate and incubated to maturity at 55°C. The screening protocol is described in the text.

Submerged Culture - The medium used for the growth of submerged cultures for enzyme synthesis was either that of Cook or Lennox basic

broth with 0 - 0.5% dextrose and 125 ug streptomycin (minimum)/mL medium when required.[3,4] The growth vessels included 2 mL microfuge tubes, standard glass culture tubes, 300 mL culture flasks, 2800 mL Fernbach flasks and 14 liter stirred tank reactor (fermentor) with working volumes of 1.25 mL, 7 mL, 50 mL, 500 mL, and 10 liter, respectively.

Harvest and Pre-assay Processing - Ten liter cultures were harvested by cartridge flow filtration and centrifugation. For cultures of less volume, centrifugation at 12,000 rpm was used. The cell suspensions were washed free of medium by 1-3 centrifugations using normal saline without Mn++, or with bis-tris propane - KCl - NaCl buffer, pH 7.15, with Mn++ (standard buffer).[5] Final pellets were resuspended in the standard buffer and immediately frozen to -30°C. Cells for dry weight measurements were most commonly made from nonwashed centrifugation pellets after drying at 80°C for 24 or more hours.

Enzyme Activity Estimations - OPA anhydrase activity estimations were made of washed whole cells ruptured by either 1 or 6 cycles of freezing and thawing in standard buffer.[5] The assay reaction mixture consisted of 5 uL of ruptured cell suspension and 47 uL of standard buffer made 0.2 mM with respect to 4-nitrophenyl methyl(phenyl)phosphinate (MPP).[6] The assay temperature was set at or about 28°C and optical density read at 400 or 405 nm. The extinction coefficient of p-nitrophenol used in computing activity was at pH 7.15, 11.85×10^3.[7] Activity is expressed as uM of substrate hydrolyzed/min·mg dry weight^{-1} or mL^{-1} culture at 28°C. Spectrophotometric readings were made either with LKB Ultrospec spectrophotometer or a Molecular Devices Thermomax microplate reader and the reaction mixture volumes used with each instrument were 2600 uL/tube and 260 uL/well, respectively.

Results

In 1984 three thermophilic strains (JD100, JD200 and JD300) were isolated at U. S. Army Chemical Research, Development and Engineering Center, whose whole cells had low hydrolytic activity against DFP (personal communication J. De Frank, CRDEC). Upon further investigation the isolates were found to be different strains of Bacillus stearothermophilus. Table 1 shows the major species-determining traits, and dissimilar strain-differentiating traits. The strains remain as originally designated JD100, JD200 and JD300. Fifty-two different traits were determined for these three different strains of B. stearothermophilus.

Strain JD100 was then mutagenized for the purpose of inducing two antibiotic markers on the chromosomal genome, resistance to streptomycin and resistance to ampicillin. Antibiotic markers are useful in growth and genetic studies as an aid in maintaining strain purity and for use as a parental stock for mutation studies. In addition it was determined earlier that the development of resistance to streptomycin in B. stearothermophilus resulted in curing the bacterium

Table 1. Diagnostic Taxonomic Traits of Bacillus stearothermophilus strains JD100, JD200 and JD300

TAXONOMIC TRAIT	JD100	JD200	JD300
Respiration			
Obligate Aerobic	+	+	+
Endospore formation	+	+	+
Gram Strain	+	+	+
Rod Morphology	+	+	+
Growth Limits, ° C			
Maximum Temperature	<73	<75	<75
Minimum Temperature	40	40	35
Indole Production	+	-	-
D Xylose Oxidation	+	-	-
Nitrate Reductase	-	-	+
Growth with 5% NaCl	-	-	+
Gas Production			
Arabinose	-	+	+
L Xylose	-	-	+
Lactose	-	+	+
β-galactosidase	-	+	-

of plasmids.[6] In this study several streptomycin-resistant strains of JD100 were induced by UV mutagenesis. However one strain, I5D, arose spontaneously and was chosen as the strain for further use because of the lack of any UV treatment. Mutants with resistance to ampicillin could not be obtained via UV irradiation. Figure 1 shows the geneology of strE I5D and the derived hyperproducing OPA anhydrase strE mutants. The mutagenized spores were spread on agar medium plates with streptomycin at a target concentration of 100 colonies per plate and incubated at 55°C for 24 to 36 hours or whenever the plate seemed ready for accurate replica plating. The protocol for primary plate

FIGURE 1. GENEOLOGY OF THE ENHANCED OPA ANHYDRASE MUTANTS OF BACILLUS STEAROTHERMOPHILUS

PARENT STRAIN
JD100

Spontaneous

STREPTOMYCIN RESISTANT
MUTANT I5D

UV Induced

ENHANCED OPA ANHYDRASE
PM MUTANTS
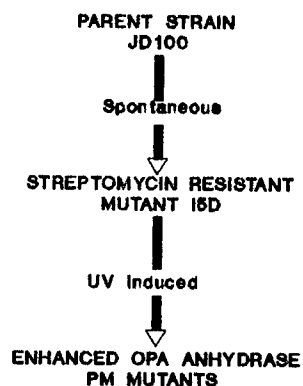
screening for enhanced OPA anhydrase strains is shown and described in Figure 2. Approximately 14,000 colonies were screened in this manner. Initially replicate filter lifts were individually tested with the 4-nitrophenyl methyl(phenyl)phosphinate, 4-nitrophenyl ethyl(phenyl)phosphinate and 4-nitrophenyl isopropyl(phenyl)phosphinate. Use of 4-nitrophenyl methyl(phenyl)phosphinate was discontinued in the plate screening because of the very rapid hydrolysis (color development time) rate at room temperature, which would make hyperproducers and nonhyperproducers indistinguishable. The bioassay overlays of 4-nitrophenyl isopropyl(phenyl)phosphinate had to be incubated at 37°C to 45°C in order to be developed but the development was poor in most cases. The 4-nitrophenyl ester of ethyl(phenyl)phosphinate was most suitable for the plate screening and its overlays generally developed within a few minutes. The substrate enzyme hydrolysis/color development controls were replicate bacteria-charged disk lifts of the experimental ones, which were denatured by heat sterilization temperatures. The selection process was strictly qualitative in nature and color development was compared within and among the combination development overlays. Three main phenotypes were selected for further testing: those that developed a strong yellow color rapidly, those that showed strong color development after a period of time, and those which showed no color development. Initially colonies that exhibited good agreement in color development on the different esters of the phosphinates were selected for further testing but this was not a prerequisite to selection for additional screening.

The primary plate screening was followed by the secondary plate screening, which involved the same plate bioassay except that eight putative mutants were grown on a plate with I5D control. Using the same qualitative process we selected 75 putative hyperproducing OPA anhydrase mutants for submerged culture testing for OPA anhydrase synthesis. The primary plate screening was discontinued after these putative mutants were obtained. A few strains found to lack activity were spurious and no negative strains were found.

The putative mutants were then tested for enhanced synthesis by growth in shaken culture in tubes containing 7 mL of medium at 55°C. After 18 to 24 hours and 44 to 48 hours the cultures were harvested for enzyme synthesis and for dry weights measurements. The amount of OPA anhydrase synthesized was computed against the dry weight of the sample. Twelve strains which showed a 25% or higher enzyme synthesis than I5D were again tested in culture tubes but samples were measured after 7 and 48 hours. In the final testing in submerged culture 8 enhanced mutants were tested for synthesis after 12, 18 and 36 hours. Figure 3 shows the putative mutants with enhanced synthesis 50% or higher than that of the control strains (JD100 and I5D). The other 4 mutants included in this testing showed enhanced activities greater than 25% but less than 50% over the controls (data not shown).

The production of OPA anhydrase synthesis by the parental strain I5D

FIGURE 2

AGAR PLATE SCREENING PROTOCOL USED TO IDENTIFY PUTATIVE
MUTANTS/CLONES WITH ENHANCED SYNTHESIS OF
INTRACELLULAR OPA ANHYDRASE ENZYME
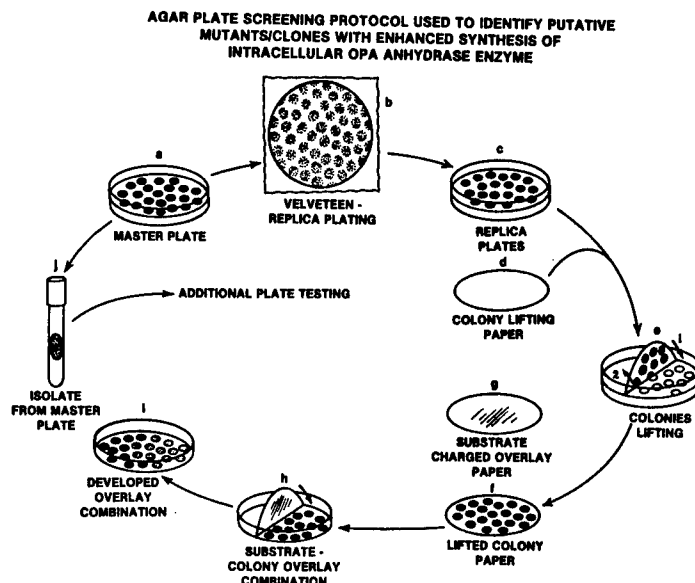
Explanation of Figure 2. a) About 100 viable, mutated bacteria are spread
on a master plate and incubated until mature; b) 2 - 5 agar medium plates
are inoculated with surface colonies from the master plate using short nap
velveteen; and master plate is reincubated until colonies imaged and then
stored at refrigerator temperatures; c) the replica plates are incubated
for growth; d) heat-sterilized filter paper disks are cut to loosely fit
into the lower plate; e) when mature, the colonies on the agar surface of
the replicate are overlaid (1) with the prepared lift paper to which we
apply gentle pressure to insure good contact between paper and colonies;
after the disk is wetted from plate moisture it is carefully removed by
lifting (2) and f) placed in a sterile petri plate and rapidly frozen to at
least -20°C; g) sized paper disks are charged with substrate by solution
impregnation and allowed to dry; h) the frozen test colony disk is placed
in a petri plate, colonies face up, and moistened with a fine mist of
buffer until saturated and then overlaid with a substrate-impregnated disk,
moistened again and the disk-sandwiched colonies incubated for catalysis
and color development at room temperature; i) during incubation the plates
are observed periodically and noted for color development; j) bacterial
isolates of promising reactions are transferred from the corresponding
colonies on the master plate to agar slant and/or another master plate and
incubated for growth and continued testing.

FIGURE 3

COMPARATIVE OPA ANHYDRASE SYNTHESIS OF PARENTAL AND FOUR PM ENHANCED MUTANTS
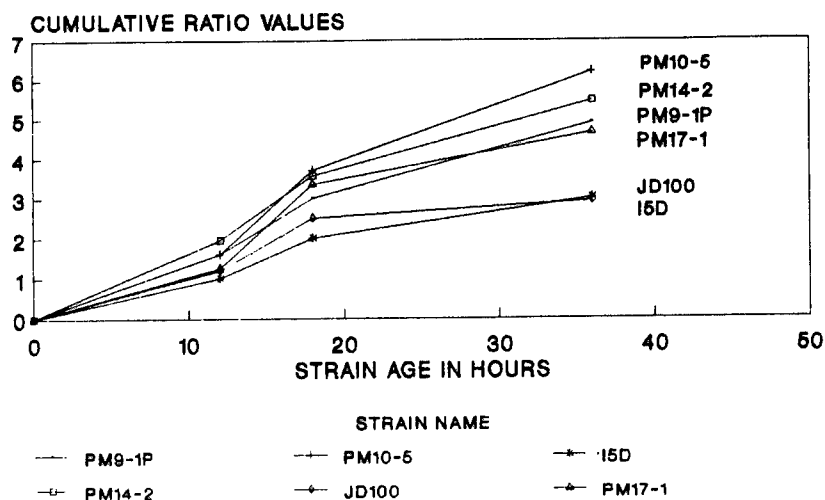
Explanation of Figure 3.
Graph showing the respective cumulative ratio values of 6 strains of B. stearothermophilus where each point represents the ratio of OPA anhydrase activity of tested strain :: OPA anhydrase activity of I5D in OD of nitrophenol formed per min for cultures harvested after shaken incubation at 55°C. Each point represents the average duplicate assay value of cultures grown in triplicate. Parent strain B. stearothermophilus JD100 is also included for comparative purposes. The cumulative ratio values assigned to the control I5D are 1, 2 and 3, corresponding to 12, 18 and 36 hour cultures, respectively. The four other putative enhanced mutants are not shown but they all have a ratio at 36 hours that lies below that of PM17-1 and a value 25% greater than that of the standard strain I5D.

---

was scaled up to 10 liters by growing it in a stirred tank reactor and under constant temperature of 55°C and the pH was controlled at 6.8 with 1N KOH. Figure 4 shows the growth kinetics and OPA anhydrase synthesis during this fermentation. The fermentation was stopped upon the rise in pH to 7. The OPA anhydrase synthesis per gram dry weight follows the same pattern as that based on turbidity in Klett Units, an increase in OPA anhydrase synthesis with an increase in dry weight. The fermentation had an approximate OPA anhydrase productivity of 0.45 units per liter per hour. This fermentation and the postharvest measurement of activity were not optimized and represent preliminary yields and productivities.

FIGURE 4

GROWTH AND OPA ANHYDRASE SYNTHESIS
BY BACILLUS STEAROTHERMOPHILUS I5D
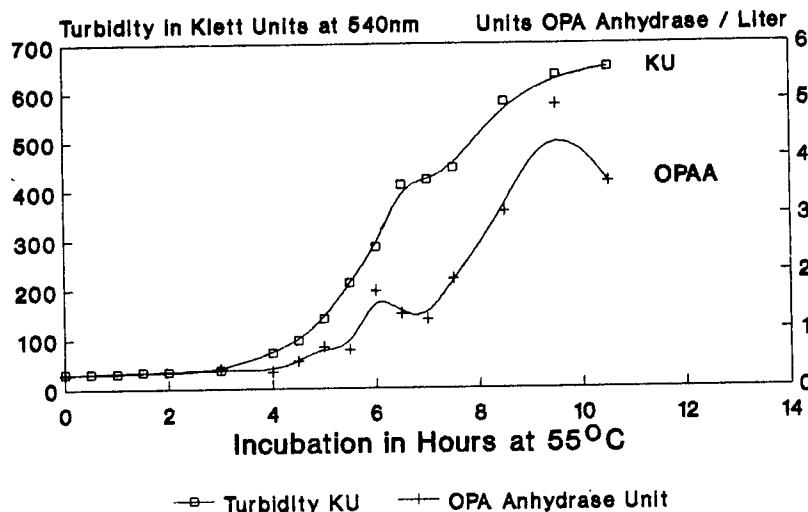
Explanation of Figure 4 - This figure describes the growth of $\underline{B}$. stearothermophilus in a 10 liter fermentation and accompanying OPA anhydrase synthesis. Each liter of LB broth was supplemented with 5 g glucose and 250 mg of streptomycin. A 10% inoculum with an age of 12 hours was used to seed the fermentation. The pH was not allowed to drop lower than pH 6.8 by 1N KOH titration. The culture was sparged with air at 5 L/min, agitated at 250 rpm and had a head pressure of 5 psi. Growth temperatures were controlled at 55°C. Washed, once frozen-thawed cells were used in estimating enzyme activity expressed as uM of methyl-O-nitrophenyl(phenyl)phosphinate hydrolyzed/min·L culture^{-1}

Discussion and Conclusions

This study describes an effective but qualitative protocol for screening for mutants/strains with enhanced synthesis of intracellular OPA anhydrase. The protocol's use is not restricted to strains of the genus Bacillus and has been used with other bacteria including strains of Escherichia and Flavobacterium. The protocol should be applicable to other intracellular enzymes/proteins synthesized in cells whose activity or presence can be detected through some color-forming product, for example, intracellular protein. This simple protocol should also be usable with any colony-forming bacteria possessing moist, sticky surface that will adhere to filter paper. The process easily lends itself for use in robotization and computerization for large-scale screening for endogenous enzymes. The protocol also allows the testing of bacterial

strains with different growth and substrate requirements on the same paper through the use of replica plating and multiple plate colony lifting.

Several forms of evidence indicate that the gene for this OPA anhydrase in B. stearothermophilus I5D has a chromosomal locus and not an extrachromosomal one. Preliminary but not exhaustive plasmid screening of its parental strain JD100 did not provide positive results for the presence of plasmids (unpublished results). Other evidence comes from the unsuccessful attempts to induce ampicillin resistance in this thermophile although spontaneous and UV-induced streptomycin resistance is obtained with relative ease. Streptomycin resistance in Bacillus can have a chromosomal origin.[9] It has been reported that induced streptomycin resistance in B. stearothermophilus clears the species of plasmids. The gene for a similar OPA anhydrase from unidentified bacterial halophilic strains was also reported to be chromosomal because of the failure to find any plasmids.[10]

Although this OPA anhydrase is unique in its properties, other similar OPA anhydrases are quite common among living things but their function and exact location in the cell remain unverified. No OPA anhydrase negative mutants were found, a fact that suggests that this OPA anhydrase plays an essential role in B. stearothermophilus. Its synthesis in B. stearothermophilus is associated with rapidly growing cultures (see Figure 4) and suggests a role in DNA replication or in protein synthesis. What further complicates the picture is that I5D is a spore former, and no attempt has yet been made to associate enzyme synthesis with sporulation.

Since the number of submerged culture samples that required testing was so large, the bacterial cells were ruptured by simple freezing to −30°C and thawing instead of using sound irradiation. The number of freeze and thaw cycles used in each study was either 6 or 1. Some studies show a decay of activity with repeated cycles of freezing and thawing. Even after mutant development, enzyme synthesis optimization and sound irradiation cell rupture pretreatment, the amount of OPA anhydrase synthesis is small as compared to other enzyme yields from other industrial microbes[11]. In order to continue to reduce the cost of producing this unique and militarily important enzyme it must be available in larger amounts and be made more easily recoverable than at present. To accomplish this goal the gene from strain I5D is also being cloned and inserted into multicopy expression vectors for use in extracellular enzyme production.

REFERENCES

1. Chettur, G., J. J. DeFrank, B. J. Gallo, F. C. G. Hoskin, S. Mainer, F. M. Robbins, K. E. Steinmann, and J. E. Walker. "Soman Hydrolyzing and Detoxifying Properties of an Enzyme from a Thermophilic Bacterium". Fundam. Appl. Toxicol. 11, pp. 373-380, 1988.

2. Hoskin, F. C. G., G. Chettur, S. Mainer, K. E. Steinmann, J. J. DeFrank, B. J. Gallo, F. M. Robbins, and J. E. Walker. "Soman Hydrolysis and Detoxication Properties by a Thermophilic Bacterial Enzyme" in Enzymes Hydrolyzing Organophosphorus Compounds, Chapter 5, Ellis Horwood Publishing Ltd., Chichester, Eng., 1989.

3. Cook, A.M. and Brown, M. R. W. "The Relationship Between Heat Activation and Colony Formation for the Spores of Bacillus stearothermophilus". J. Pharm. Pharmacol. 16, 725-732, 1964.

4. Lennox, E. S., "Transduction of Linked Genetic Characters of the Host by Bacteriophage P1". Virology 1, 190, 1955.

5. Hoskin, F. C. G., and Roush, A. H. "Hydrolysis of Nerve Gas by Squid Type Diisopropylphosphorofluoridate Hydrolyzing Enzyme on Agarose Resin. Science 215, 1255-1257, 1982.

6. Lieske, C. N., J. H. Clark, H. G. Meyer, M. A. Lawson, J. R. Lowe, P. Blumbergs, and M. A. Priest, "Inhibition of Two Acetylcholinesterases by the 4-Nitrophenyl Esters of Methyl-, Ethyl-, and Isopropyl(phenyl) phosphinic Acid. Pest. Biochem. Physiol. 17, 142-148, 1982.

7. Martin, C. C., J. Golubow, and A. E. Axelrod. "A Rapid and Sensitive Spectrophotometric Method for the Assay of Chymotrypsin". J. Biol. Chem. 234, 294-298, 1959.

8. Imanaka, T., M. Fujii, I. Amamori, I., and S. Aiba, "Transformation of Bacillus stearothermophilus with plasmid and characterization of shuttle vector plasmids Between Bacillus stearothermophilus and Bacillus subtilis". J. Bacteriol. 149, 824-830, 1982.

9. Staal, S. P. and J. A. Hoch, "Conditional Dihydrostreptomycin Resistance in Bacillus subtilis", J. Bacteriol. 110, 202-207, 1972.

10. Cheng, T. C., and J. J. DeFrank, "Screening for Plasmids in Halophiles". Technical Paper CRDEC-TR-094, 13 pp., US Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD. 1989.

11. U.S. Patent No. 4,472,504, September 18, 1984. <u>Benedict J. Gallo</u> Inventor. "Cellulase-Producing Microorganism".

Hybrid Approaches for Spread-Spectrum COMINT
C. Garvin, D. McGuire, M. Younger
U.S. Army Harry Diamond Laboratories
2800 Powder Mill Rd
Adelphi, MD 20783
J. Deberry
Air Force Institute of Technology
WPAFB, Dayton, OH 45433

Introduction

The difficulty of gathering communications intelligence (COMINT) is exacerbated by the use of spread-spectrum techniques. Noncooperative determination of the bandwidth spreading scheme used in any specific case is a difficult processing problem. In a typical tactical communication-band scenario, the listener must search the band of interest (say the HF, VHF, and UHF bands) to detect radiant energy, ascertain the modulation scheme, and properly demodulate the information before the signal goes off the air. Message lengths vary from seconds to several minutes; however, frequency-hopping (FH) systems typically change carrier frequency hundreds or thousands of times during each second. A simple calculation illustrates the difficulty in predicting or following a frequency-agile COMM signal rapidly enough to exploit the information being sent.

Optical COMINT Processors

Many optical processors are designed for use as wideband intercept receivers: acousto-optical (AO) spectrum analyzers [1] and correlators (pulse compressors) [2] are two of the most successful in signal detection and analysis. These optical processors offer wide instantaneous bandwidths, limited chiefly by the AO interaction which places the signals of interest on an optical carrier. Bragg cells which use the bulk or surface AO interaction allow up to several gigahertz of bandwidth [3].

Typical signal-acquisition systems employ a hybrid approach [4], using a wideband optical system (spectrum analyzer or correlator) to detect the energy, analog

rf filtering to remove interfering signals, and digital processing elements to track and determine modulations. By applying the best suited technology to solve particular barriers of signal acquisition and analysis, one can build hybrid systems that can extract signal features of low probability of intercept (LPI) signals. We conducted experiments using two such hybrid intercept systems consisting of an optical processor (spectrum analyzer and one-dimensional correlator), analog-to-digital (A/D) board, digital hardware, and MC68010 microprocessor. The data acquisition system is shown in figure 1.

A simulator was developed which is capable of several forms of spread-spectrum modulation. The simulator drives a PTS synthesizer (which has a rapid switching speed of about 10 μs) either linearly stepping or randomly hopping in frequency. A direct sequence (DS) modulator was used to spread the signal from the PTS synthesizer. In these ways the various forms of spread-spectrum signal were simulated. In the experiments reported here data were not encoded on the signal in any form (such as AM, FM, or binary phase shift keyed); however, previous experiments have demonstrated the independence of the intercept methods used here to these modulation schemes. We first discuss experiments conducted using an AO spectrum analyzer as the signal sensor.
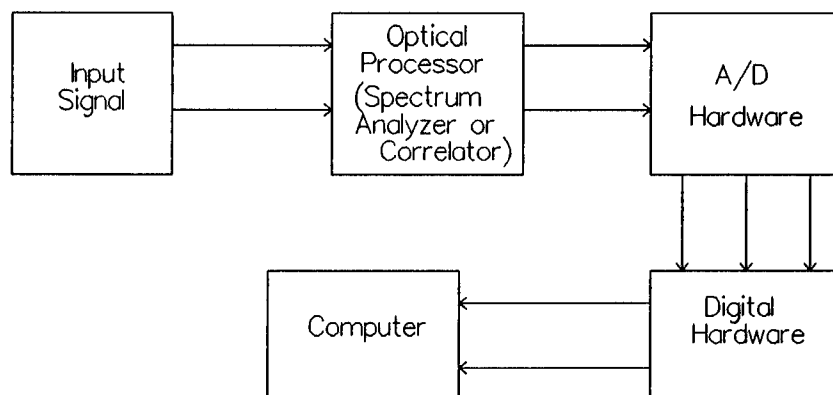
Figure .1: Acquisition hardware

Wideband Spectrum Analysis

The AO spectrum analyzer uses the linear relationship between the frequency of the acoustic wave and the Doppler shift of the optical wave it refracts in the AO interaction. A lens is placed in the optical path after the AO Bragg cell, and since the lens refracts the light proportionally to its frequency, differently Doppler shifted light is spatially separated. A detector array placed at the focus of the lens detects

this spatially separated light forming the power spectral density of the signals on the AO Bragg cell.

An AO spectrum analyzer designed for the tactical HF radio spectrum was used to generate the data shown in figure 2. Although the optical spectrum analyzer gives 60 MHz of instantaneous bandwidth (at a 30-kHz resolution), the digital hardware allowed only 12 MHz of usable bandwidth. The graph shows the small section of the frequency-time plane that the digital hardware can handle in real time. The "pedestals" in the graph represent two signals that moved asynchronously into and out of this window. Figure 3 shows the results of a least-squares centroid algorithm operating on these data. The centroider was implemented on a VAX 11/780. Algorithms have been developed to use such centroided data to follow signals at high hopping rates. Other curve-fitting routines for centroiding were implemented; for example, a sinc-squared routine was used with FH-DS hybrid signals in order to obtain simultaneous estimates of bandwidth and center frequency for each hop.
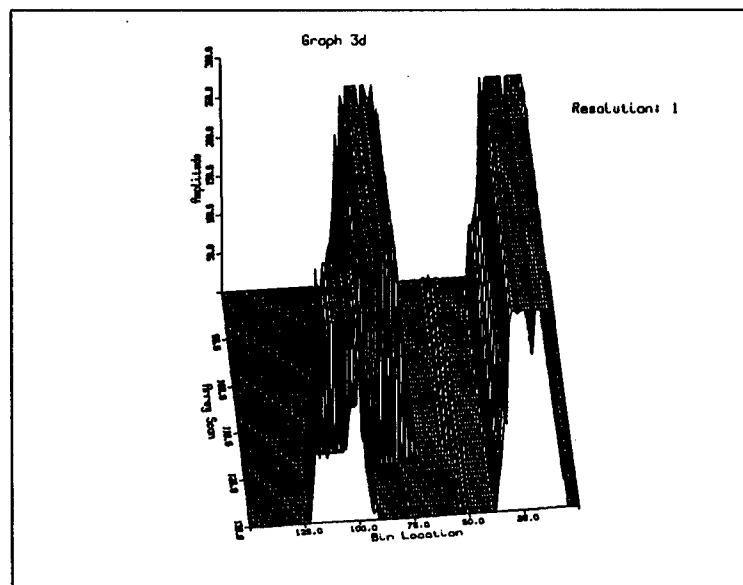
Figure .2: Spectrum analyzer output

Advantages Using Autocorrelation

Some spread-spectrum signals will go undetected or elude evaluation without a receiver matched to their signature. This is particularly true if the power in the
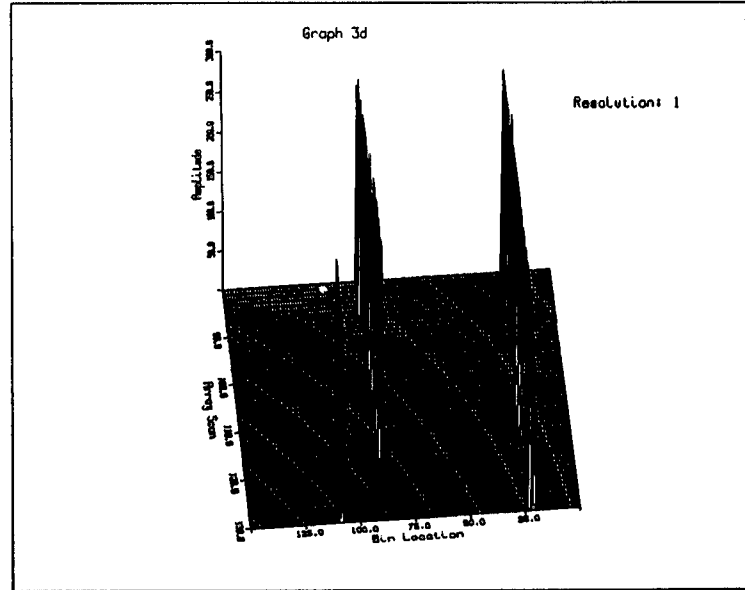
Figure .3: Centroider results

transmitted signal is low in comparison with channel noise. Optical correlators offer a unique capability to receive and analyze those signals in the presence of noise. If the power of a signal is spread out over the frequency spectrum, then the signal will be difficult to detect by conventional methods. By correlating in the time dimension, we take advantage of several characteristics of correlation. Since noise on independent channels of the correlator does not correlate well, a correlator can be used to improve the signal-to-noise ratio (SNR) of a received signal. This method (autocorrelation) can reduce noise, thus giving a high processing gain with no *a priori* knowledge of the signal. Certain spread-spectrum signals display unique correlation properties, allowing some characteristics of their autocorrelation (such as shape, size, and position) to be used as sorting parameters.

Different versions of the above-mentioned spread-spectrum signals were constructed using a hybrid signal generator and fed directly to the correlator. Plots of the digitized correlator output (shown in figures 4, 5, and 6) indicate clear differences among the correlation signatures of the different spreading schemes.

The type of spreading used, as well as other signal externals such as bandwidth, center frequency, and time of arrival, can be extracted from the raw correlations or the fast Fourier transforms (FFT's) of these correlations. By performing a digital Fourier transform on the correlated signal, we obtain the power spectral density (PSD) of the signal. For the purposes of this review, we consider examples of such signals having a positive signal-to-noise ratio. The example signals discussed
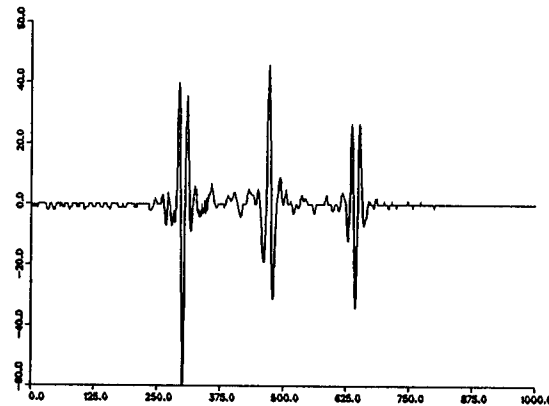
Figure .4: Correlation of linear frequency stepper

below were presented to the hardware in a benign environment: signal and channel noise were present, but no interfering signals are included. Signal internals are not discussed in this paper.
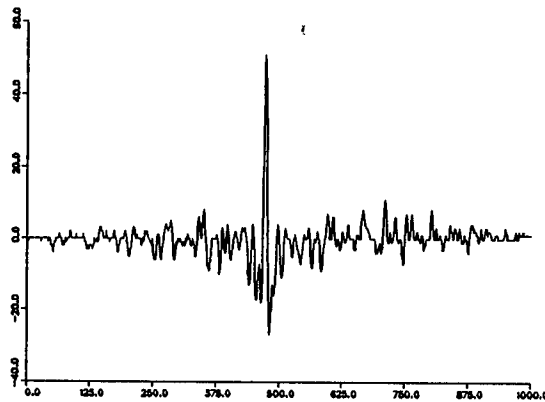
Figure .5: Correlation of random frequency hopper

Studies were conducted on the autocorrelation of signals having different modulations using various lengths of time integration in the correlator. These experiments demonstrated that correlation signature can be used to distinguish among the different forms of spread-spectrum. A typical FH signal will have a power spectral density with a pedestal-like appearance. The width of the "pedestal" represents the bandwidth over which the signal was "hopped." The envelope of the FFT of the correlator output of a DS signal is a sinc squared function whose centerlobe width (zero crossing to zero crossing) is equal to twice the code rate. The power spectral density of a combination of an FH and a DS signal — one class of hybrid — has an envelope which is the product of these two, or more simply, power is distributed

31
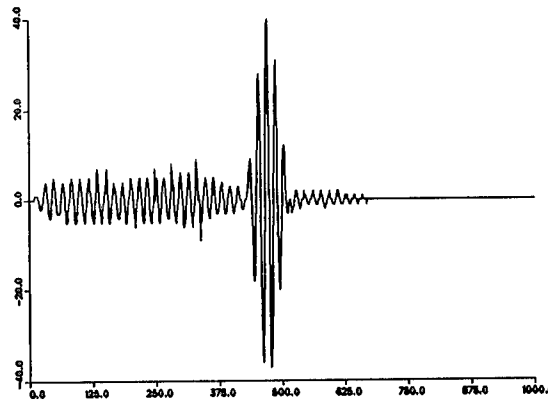
Figure .6: Correlation of direct sequence spread

equally over the hopped bandwidth and the PSD of this signal appears as a flat top with sinc squared edges.
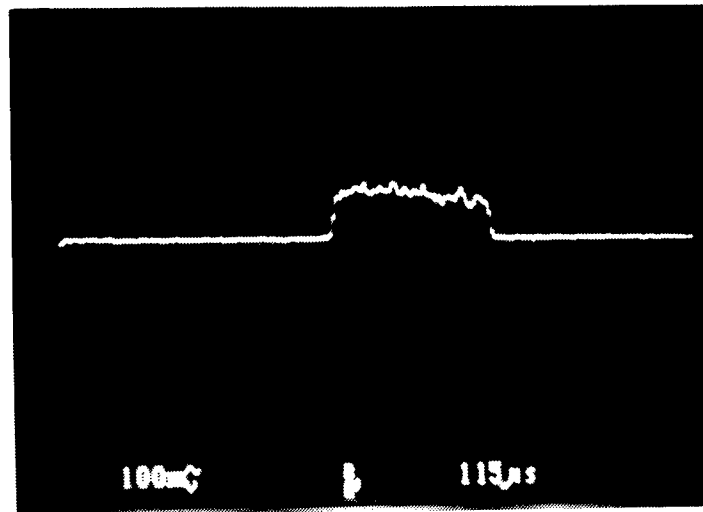
Figure .7: PSD of typical FH

Correlator operation requires the adjustment of integration time, system IF (to bring the signal into correlator bandwidth), and several other system parameters. In the selection of a proper integration time, several factors must be considered. Since the signals of interest are likely to be moving around in frequency at a high rate, is there an optimal value for the correlator integration time (τ) ? Clearly, the frequency resolution available from the FFT of the correlation is limited (primarily) by the correlator's time-bandwidth product, or simply the number of resolvable spots. The time resolution is limited to the integration-and-readout time of the out-
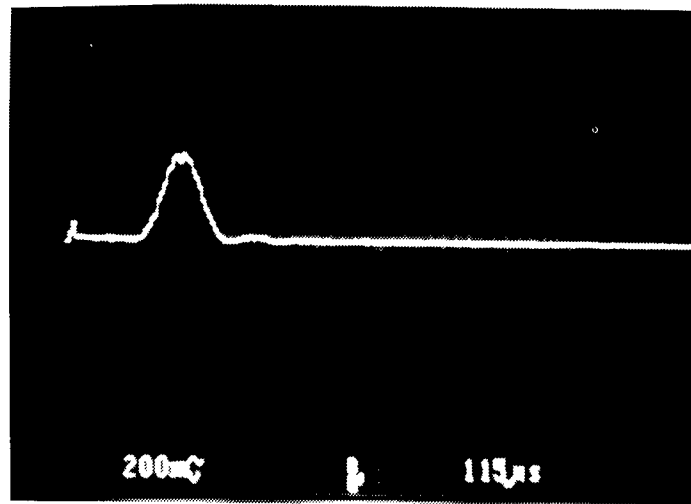
Figure .8: PSD of typical DS

put detector of the correlator, which in this case is a linear self-scanned photodiode array. The fastest hopper you can hope to follow would be moving at a rate of $\frac{1}{\tau}$ although you can in principle detect frequency-agile signals having faster rates by their foldover correlations (an example of these foldovers is seen in figure 4).

Neural Postprocessing

Currently, the AO processor output is handled in one of two ways: the detector output signal can be evaluated by an operator directly, or the signal can be digitized and passed through curve-fitting routines in a digital computer; in both cases the goal is to recognize signatures and thereby extract information about the signal(s) present. It was postulated that an artificial neural system (or neural net) could be used to learn correlation signatures and to classify signals by means of their correlation. The neural net could be trained to recognize signals having similar correlation shapes, although they may have different bandwidths and/or amplitudes. Another net could be used to classify bandwidths, and a third to classify correlation amplitudes. With the appropriate learning algorithms, these nets could be set up to train themselves on new correlation signatures and catalog new signatures independently. Experiments to this end were performed in collaboration with the Air Force Institute of Technology.

Several AO correlators could be used to provide data, each of which has a slightly different data format. Correlator output can be provided as vectors of 256 elements, 1024 elements, or 4096 elements long whose elements take on values ranging from 8
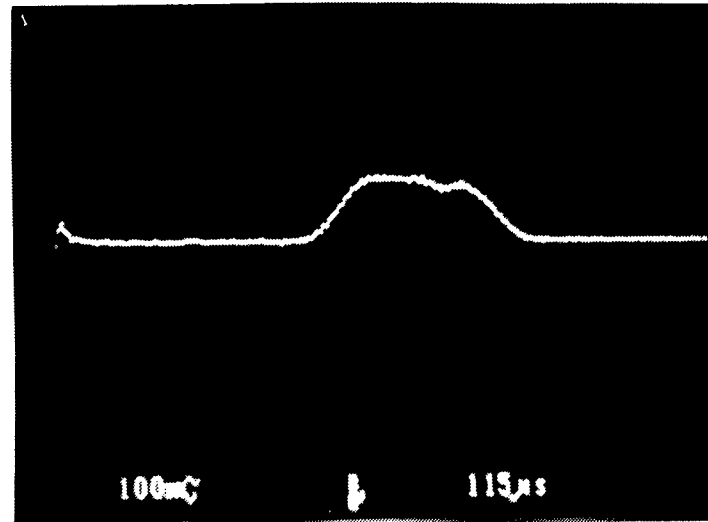
Figure .9: PSD of FH-DS hybrid

bits to 16 bits. Some correlators can provide these vectors as rapidly as one every 10 μs, although the update rate depends on the correlator integration time and we may integrate some vectors for as long as seconds. Although it would be possible to program the neural net for various rates and send a clocking signal, it was decided that the feasibility study would not require such generality; therefore we selected data from a one-dimensional correlator whose 1024-element detector updated every 10 ms. Training data for the net consisted of about a hundred vectors for each class of signal (see figures 4, 5, and 6 for examples of raw correlations), representing correlations of the same sort of signal at several different bandwidths, center frequencies, code rates, and hopping rates. Two classes of signal were presented to the net; we chose to have the net report the correlation shape (and/or the signal structure which it represents). Because of limitations in computing power, it was impractical to operate the net with the full 1024-element vector. For input data reduction, some limited filtering was performed on these correlation signatures. For each signature every two data points were averaged, this result was normalized, and the central 50 datapoints were fed to the network.

A three-layer perceptron (shown in figure 10) was coded on a SUN4 workstation and trained with supervision on 30 versions of the datasets described above. As each training exemplar was presented to the net, the correct class was compared to the network output. Error signals were propagated back iteratively through the hidden layers of the network; a back-propagation learning algorithm was used to adjust the initially random interconnection weights. Training vectors were iterated for 20,000 cycles; usually no improvement in performance was observed for longer
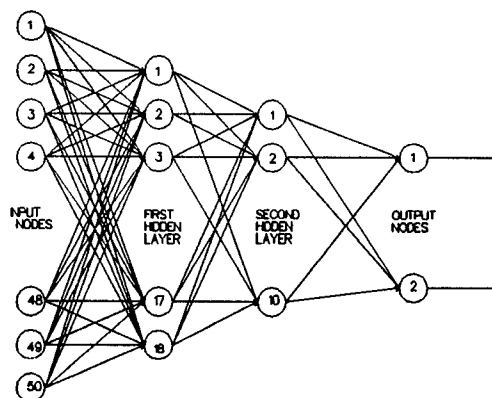
Figure .10: Three-layer back-propagation net used in AFIT study

training cycles. In an additional experiment, parallel nets were trained in groups of three on partially disjoint training sets; the same test vectors (all different from the training vectors) were presented to these nets and their outputs were combined using a majority vote decision rule.

From the results of these experiments, the three-layer back-propagation network was shown to successfully classify wideband correlation signatures. The results show that such a net can be expected to correctly classify 100% of training correlations. It can be expected to correctly classify at least 80% of untrained correlations. This compares well to the degree of success expected from a linear discriminant function. A performance improvement was observed when multiple parallel nets were used on the same test correlations only if the training sets of parallel nets differed. An improvement of 2% in classification accuracy was observed when the training sets differed by 17%. Classification performance varied with variations in initial connection weights, training sets, and order of training set presentation. The most significant of these was the training set composition.

It is our intent to extend these results to a four-class correlation recognition problem. We would like to test for further classification improvement using the parallel nets and majority-vote decision rule. We hope to develop hybrid digital/optical hardware to implement the neural recognition algorithms demonstrated here. Coupled with recent minibench optical correlators [5], such a system could provide processors powerful and compact enough for applications on unmanned aerial vehicles, anti-radiation missles, or tethered platforms (figure 11).
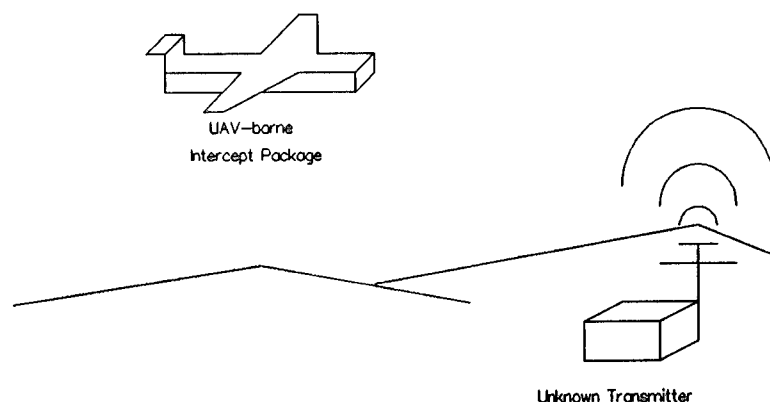
Figure .11: UAV intercept scenario

Summary Discussions

Individually, FH and DS signal structures are very challenging targets for noncooperative reception/exploitation. In combination they can easily elude conventional intercept receivers. The approaches discussed above offer the capability of real-time detection and processing of hybrid FH-DS signals.

In the preceding sections we discussed only two of the myriad optical processing architectures useful in signal-processing applications. Although there are now channelized receiver systems using fast A/D's and digital processing for some applications, none can achieve the instantaneous bandwidth with the size/weight/power efficiency available with standard Bragg cell technology. In addition to their COMINT applications, optical correlators can be applied to a variety of high-order spectrum analysis problems, such as the real-time calculation of the bispectrum. This higher-order statistic has generated much interest [6,7] , and is actively being pursued for signal and image processing. A two-dimensional acousto-optic correlator architecture which we have constructed [8] can be used to calculate this statistic. When coupled with a versatile digital processor, the optical modules discussed above provide powerful tools for the characterization of signal intelligence broadcast by noncooperative transmitters.

From earlier discussions concerning Bragg cell components, it is easy to see that these techniques can be feasibly used in the interception and exploitation of LPI radar signals (FH radars, DS radars, and hybrids). Because modulation devices having as much as 2 to 3 GHz of bandwidth are available, optical architectures are suitable for many applications in this range. Since there are many wide-bandwidth signals (such as impulse radars, anti-personnel radars, etc.), numerous threat signals can be exploited with these optical techniques.

Bibliography

[1] M. W. Casseday, N. J. Berg, I. J. Abramovitz, A. N. Filipov, and C. Garvin, *Advances in Bragg Cell Spectrum Analysis*, Proc IEEE Ultrasonics, vol. 2, p731, October 1981.

[2] N. J. Berg, J. N. Lee, ed., **Acousto-Optic Signal Processing: Theory and Implementation**, Marcel Dekker, Inc., New York, 1983.

[3] J. C. Poncot, P. Nyeki, Ph. Defranould, and J. P. Huignard, *3 GHz Bandwidth Bragg Cells*, Proc IEEE Ultrasonics, vol. 1, p501, 1987.

[4] B. M. Sadler, G. Cooley, A. Filipov, M. Patterson, and J. Pellegrino, *Hybrid AO System for Wideband Frequency Agile Signal Identification*, Proc MILCOM '88, vol. 3, p67, 1989.

[5] R. R. Ulrich, I. J. Abramovitz, N. J. Berg, and M. W. Casseday, *A Compact Surface-Wave Acousto-Optic Time-Integrating Correlator*, Proc MILCOM '82, vol. S, 1982.

[6] G. B. Giannakis, and M. Tsantanis, *Object detection and classification using higher-order statistics*, 6th IEEE Workshop on Multidimensional Signal Processing, September, 1989.

[7] B. M. Sadler, *Detection and classification using higher-order statistics of optical matched filters*, Advances in Optical Information Processing IV (Proc SPIE 1296, D. R. Pape, ed.), SPIE OE/Aerospace Sensing, April 1990.

[8] C. Garvin, N. J. Berg, and R. Felock, Proc Ultrasonics International '85, p429, July 1985.

Optical Associative Memory Based on Binary Phase Algebra and the Inner-Product Architecture

John R. Goff and Deborah K. Shockley *
Harry Diamond Laboratories
2800 Powder Mill Rd., Adelphi, Md. 20783-1197

1 Theory

We report here the optical implementation of an associative memory with several characteristics which set it apart from other optical implementations. Specifically we discuss these characteristics: (a) this associative memory is an *inner-product* realization, (b) it makes use of the phase-controlling effects of magneto-optic spatial light modulators (MOSLM's), and (c) by virtue of thresholding at an *interior* plane, it has superior recognition features and memory capacity. Not detailed here are its hybrid nature (i.e., optical and electronic) and the microprocessor control of all the functions of the optical associative memory (OAM).

First of all, we point out the architecture employed for this associative memory (Fig. 1). Information in the form of a binary-valued vector with N components (bits) is entered, with replication, into the columns of spatial light modulator (SLM) 1. Memory vectors, of which there are M, are loaded by columns into SLM 2. A coherent collimated laser beam traverses the two SLM's, imaging the pixels of 1 perfectly onto those of 2. This optical system, which is completed by the addition of a polarization analyzer and an integrating lens, functions to perform a correlation between the input vector and each of the M memory vectors. This is done in what we may call a "phase-mode" operation. The SLM's are bipolar in the sense that the polarization of light traversing a pixel is rotated either clockwise or counterclockwise, depending on the state (bit value 0 or 1) given that pixel. Where the state of the pixel at x, y of SLM 2 agrees with that of the pixel at the same x, y coordinates of SLM 1, there is net polarization rotation of the light intersecting both; with the analyzer crossed with respect to the laser radiation, some light passes the analyzer. In contrast, where there is a disagreement and therefore cancellation of rotation, no light can pass the analyzer. In this way agreements and

disagreements between input vectors and memory vectors are realized as differences between *light* and *dark*. The extinction ratio, *light* versus *dark*, is very high when MOSLM's are used in this mode of operation. [1] The pixels for any input-memory comparison (i.e., for any column) are summed by the lens, so that there results a linear array of spots at plane C (Fig. 1), representing the M correlations, indexed by $k(1 \leq k \leq M)$.
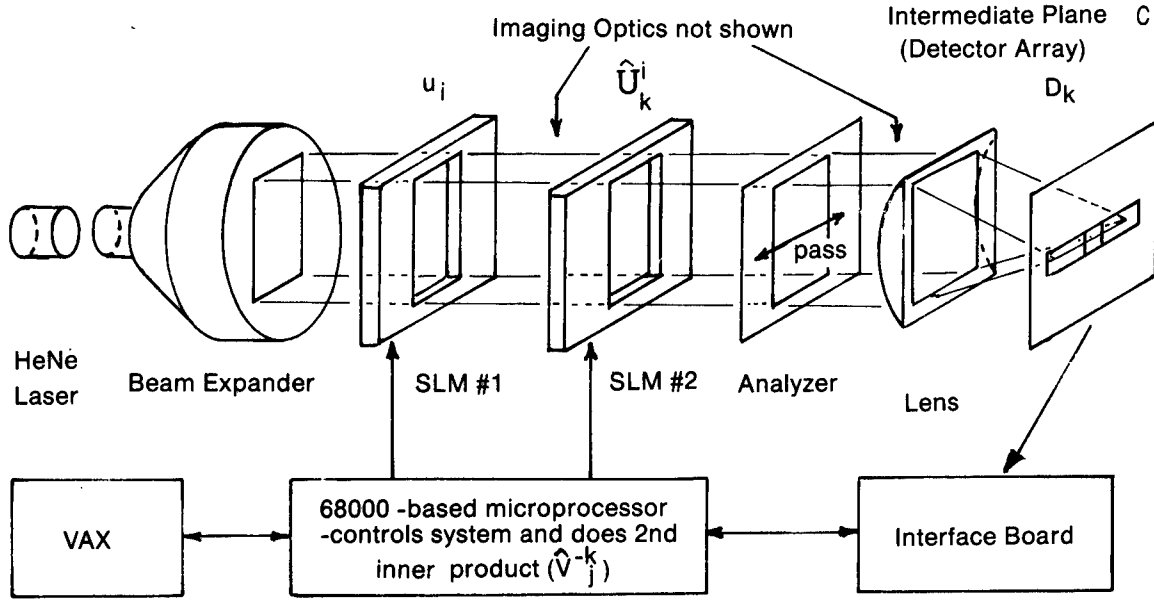
Fig. 1: Diagram of inner-product optical associative memory, constructed for this study

1.1 Mathematical Description

So far the system as described is simply a one-dimensional pattern correlator. However, as we shall see, it implements the first half of a full-blown, single-layer associative memory (AM) and has several attractive characteristics which most AM's lack. The mathematical expression for the system described above, constituting the "optical half" of Fig. 1, is

$$d_k = \sum_i \hat{u}_k^i u_i, \tag{1}$$

where u_i is the input vector (components indexed by i) and $\hat{u}_k^i$ is the k-th memory vector. Correlation is represented by an inner-product contraction. The result of this

contraction is a vector, d_k, which represents the degree of correlation for the k-th memory. To complete a full inner-product associative memory (IP AM), we need a second inner product,

$$v_j = \sum_k \hat{v}_j^k d_k,$$ (2)

where $\hat{v}_j^k$ is another array of memory vectors, which may be identical to $\hat{u}_k^i$. In the hybrid arrangement of Fig. 1, this second operation is performed digitally in an associated microprocessor.

The output of the AM is the vector v_j, which is considered the *associate* of the memory vector $\hat{u}_k^i$ most closely resembling the input u_i. Thus the name "associative memory." The system functions to correct errors: i.e., if the input is a corrupted version of some stored vector $\hat{u}_k^i$, the output will be the corrected vector in an autoassociative system — where $\hat{u}_k^i$ and $\hat{v}_j^k$ are identical. Some error-correcting capability is inherent in this AM, [2] but this capability can be greatly enhanced through thresholding at plane C, as we shall see. Some of the metrics on which the performance of the AM are to be judged are (a) the convergence length, i.e., the largest distance between the input and the memory that can be tolerated before recognition is compromised, and (b) the memory capacity, i.e., the number of memory vectors M which can be packed into a reliable AM system operating on vectors of dimension N.

An AM is not substantially different from a neural net. Consider Eq. (1) and a "network" representation of it, as in Fig. 2. The matrix elements $\hat{u}$ of Eq. (1) are the interconnection weights of the network of Fig. 2. The input nodes are the input vector components u_i. The output nodes, or neurons, are the thresholded output vector components, here d_k. The various weights correspond to "synapses" in the neural picture. In principle, the AM of Eqs. (1) and (2) can be ramified to perform most functions associated with neural networks. These include solution of optimization problems and pattern-recognition tasks.

Usually neural networks are compared with the "outer-product" representation of AM's. The outer-product (OP) representation of an AM is obtained by collapsing the inner-product (IP) equations. Substituting Eq. 1 into Eq. 2 to eliminate d_k, we obtain

$$v_j = \sum_i^N w_j^i u_i,$$ (3)

where

$$w_j^i = \sum_k^M \hat{v}_j^k \hat{u}_k^i.$$ (4)

The weights of the OP AM are formed as the outer product of the two memory matrices, and these weights w_j^i constitute the interconnections in the usual picture of a one-layer neural network. In this case v_j, after thresholding, is the neuron output.
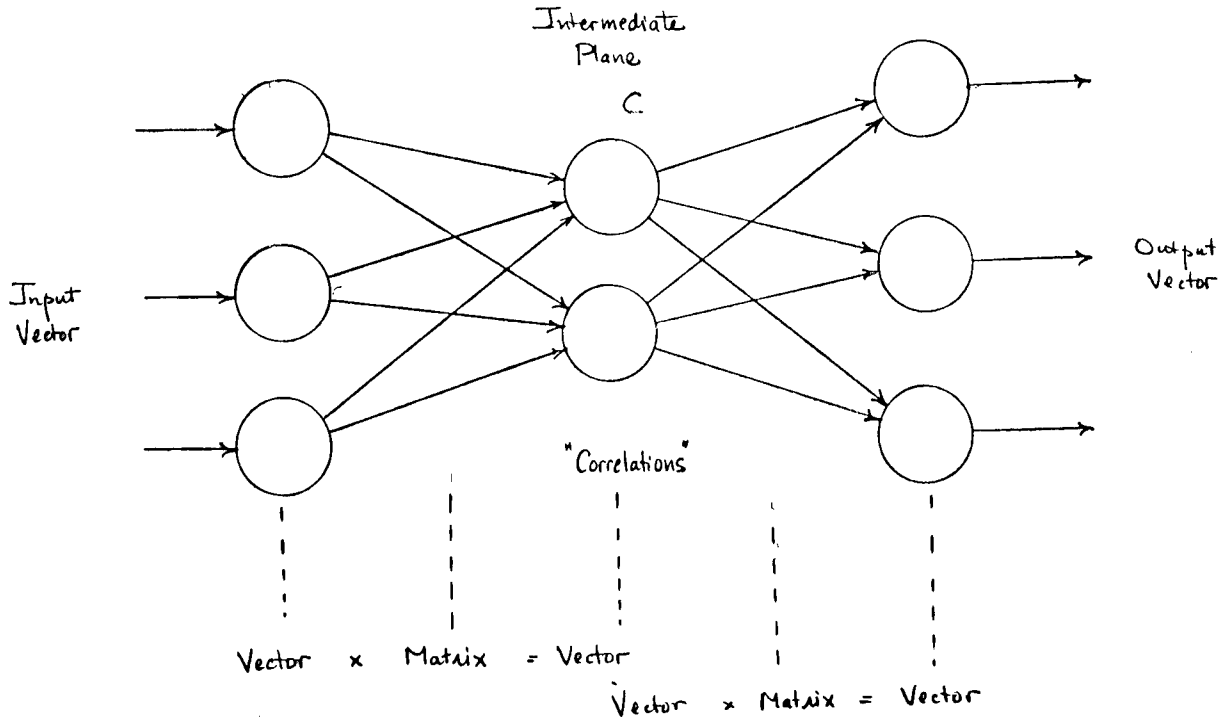
Fig. 2: Network representation of associative memory. Note correspondences of location of plane C here and in Fig. 1.

1.2 Learning

One advantage of the IP type of AM over the equivalent OP is immediately evident. In the former, the memory vectors, assumed known, are loaded into the SLM's and the optics takes care of the rest. In the latter, a new matrix w must be calculated from $\hat{u}$ and $\hat{v}$ via Eq. (4) in an off-line step. A common approach is to form w not by such a calculation but by an iterative process called *learning*. This is usually computationally intensive and such learning is usually slow. To the extent that similar results can be obtained with simple addressing of an SLM, the inner-product approach is seen to have a potential advantage in terms of speed and simplicity.

This advantage is particularly meaningful in the context of optics. In the electronic (digital) realm, back-propagation learning may be tedious, but the algorithm is straightforward. In the optical realm, there is always a strategem to do back-propagation, but it is complex, unnatural, and sensitive to optical imperfections and analog frailties. Those who wish to implement AM's with optics should look closely at the inner-product approach to see if iterative learning might be avoided for the application at hand.

We have found that, in an AM context, *iterative* processing in the association function is largely unnecessary if the vector components are binary rather than multi-valued (grey-level) or analog. In other words, if recognition is possible, it should happen on the first pass, and a further pass (after setting $u_i = sgn(v_j)$) cannot do any more. Binary information, though acceptable for digital computers, does not fit so well with optics. Nevertheless, binary data handling in the context of Fig. 1 imparts these advantages: (1) one-pass recognition and (2) accurate processing using the most common types of SLM, which are binary. These advantages may override other considerations and motivate the IP optical AM implementation.

1.3   Threshold

Perhaps the strongest point in favor of an inner-product (IP) optical implementation is the ability to threshold the variable d_k at plane C. In practice this means that, with such thresholding, a system may hold many more memory vectors and still perform recognition functions with the same effectiveness. A conventional Hopfield AM is reliable only if $M < N/(2\log N)$ in theory, or about $M = 0.15N$ in practice. [2] However we have shown that with judicious choice of threshold, a capacity of $M = N$ retains effective recognition. Thresholding serves to excise the lower values of d_k, neutralizing their ability to affect recognition adversely through the introduction of statistical noise. This is illustrated in Fig. 3, in which a d_k with high positive correlation (number 8) is accompanied by many other peaks showing random bit correlations only. The thresholding, which includes subtraction of the threshold value from all d_k, must remove at least the net negative correlation elements. If the threshold is not elevated sufficiently above that level, noted as 120 in Fig. 3, some of the peaks which do not represent correct association will cause bit flips in the correct association result, and the output v_j differs in some bits from the desired $\hat{v}_j^k$.

The situation can also be analyzed mathematically by considering the following rewriting of Eq. (2) in conjunction with Eq. (1):

$$v_j = \hat{v}_j^p \sum_i \hat{u}_p^i u_i + \sum_{k \neq p} \hat{v}_j^k \sum_i \hat{u}_k^i u_i. \tag{5}$$

Note that the first term on the right, where $k = p$, representing the "right" answer, gives $\|u_p\|^2 \hat{v}_j^p$, which is thresholded as the final step to become $\hat{v}_j^p$. The whole of the second term represents noise. A mathematical analysis of the statistics of this expression shows how the interference from this term is affected by raising the threshold. [3] The improvement in recognition is dramatic. Of course, if an input vaguely resembles several memories, it cannot be identified properly no matter what thresholding is performed, and this is as it should be.
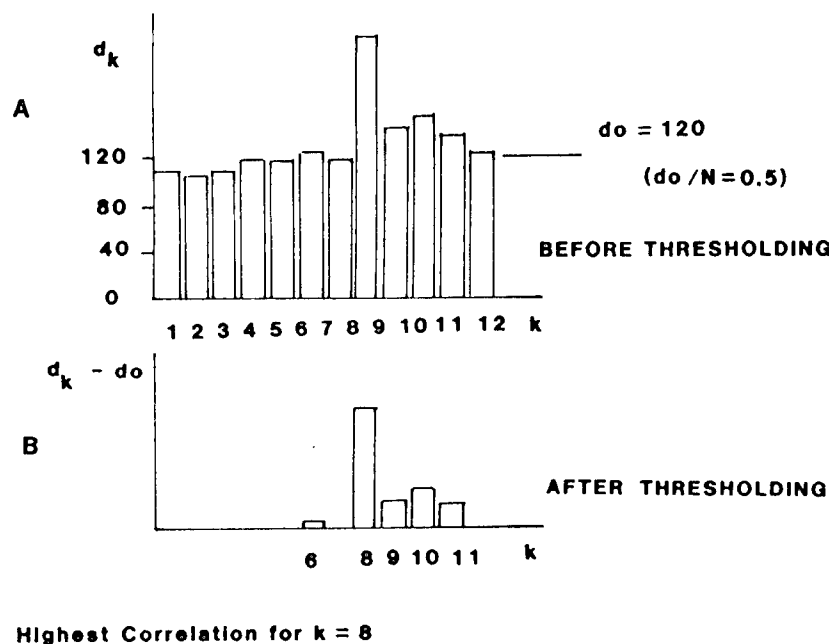
Fig. 3: Effect of threshold on recognition. (a) d_k for $k = 8$ is larger than for other k. All other d_k vary around $d_k = 120$, which corresponds to random agreements (full correlation $\doteq 240$) (input has two errors). (b) After thresholding with $d_o = 120$, all but 5 vectors are eliminated.

In the implementation of Fig. 1, the correlation results (the d_k) fall upon a linear array of *square-law* detectors, specifically a self-scanned Reticon array. The inner-product AM would then appear to be *quadratic* and, if it were, it would have superior memory capacity and related qualities. [4] In fact however, the system remains a linear

44

AM owing to the way the pixel amplitudes are coherently summed at the transform plane (plane C). [3]

1.4 Dimensional Extension

The system described here works for one-dimensional data but does not extend naturally to two-dimensional data, such as camera images. Note that the two variables of the system of Fig. 1 are spatial coordinates in a plane perpendicular to the laser propagation vector. We can add two angular variables to make four in all; this enables an IP (or OP) system to process two-dimensional data. This argument leads us to the use of angular multiplexing in holograms which would replace the (nonholographic) SLM's. It should be recognized that we are not reporting on a system built with this feature, but rather we show it as a reasonable extension of our system.

An interesting observation is that, whereas the IP approach is rather unfamiliar as a neural net approach in electronic realizations as well as in the optical processing of one-dimensional patterns, it is nevertheless almost standard in two-dimensional optical implementations. In the latter context, holographic SLM's (for instance computer-generated holograms or photorefractive crystals) are very effective, as indicated above. The IP approach fits very nicely with a holographic implementation. One reason is that reference images (or their Fourier transforms) can be introduced as such onto the holographic SLM's without the need for computing outer products as in Eq. (4). Some examples of such two-dimensional IP systems (often not identified as IP) are found elsewhere in the literature. [5-7] Generally the elements of the correlation vector d_k are realized as plane waves at various angles, as propagation through variously positioned pinholes, or as eigenmodes of a resonant cavity. Such systems are fully optical, i.e., without an electronic second inner product. This means that there must be some operation in place of the detector array at plane C, which may in part perform a thresholding function. However, the value of interior thresholding in an IP system is not generally appreciated. We feel that with an enhanced understanding of this aspect of IP processing, it should be possible to design the most effective two-dimensional pattern-recognition associative memories, increasing the memory capacity over what is presently considered practical.

While SLM's are often binary, holographic media are generally analog. Therefore, unless forced to be otherwise, the two-dimensional systems contemplated here are analog and are also generally iterative, or recirculating. We believe that iteration is necessary because of the multivalued nature of the data, although this has not been proved.

In many neural net functions, the memory vectors are not known per se. The advantage of simple loading of memories into the SLM's is of course lost in this case. It is possible however to do off-line calculations (i.e., digitally on separate apparatus) which determine $\hat{u}$ and $\hat{v}$ for Eqs. (1) and (2) from the known conditions of the problem: e.g., from known OP w_j^i weight values, from constraints of an optimization problem, or from statistical data in an unsupervised learning situation. These considerations will be addressed in another paper.

The inner-product realization of an AM appears to open up new possibilities for a simple implementation of sensor fusion. The approach is sketched for two independent data streams, u_i and x_m, deriving from the same object, as follows. The vector u_i is the processor input; the memory $\hat{u}_k^i$ is formed from the normal $\hat{u}$ memory but conditioned by the other sensor stream x_m. This conditioning is effected in a parallel AM which contains only the first half, i.e., performing the products $x_m \hat{x}_k^m$, where $\hat{x}$ is another memory array. The scheme is shown in Fig. 4. Simulations of the sensor fusion operation are encouraging, but conspicuous advantages will probably be obtained only in a noisy system with more than two sensor input streams. This application of IP processing will be reported in more detail subsequently.
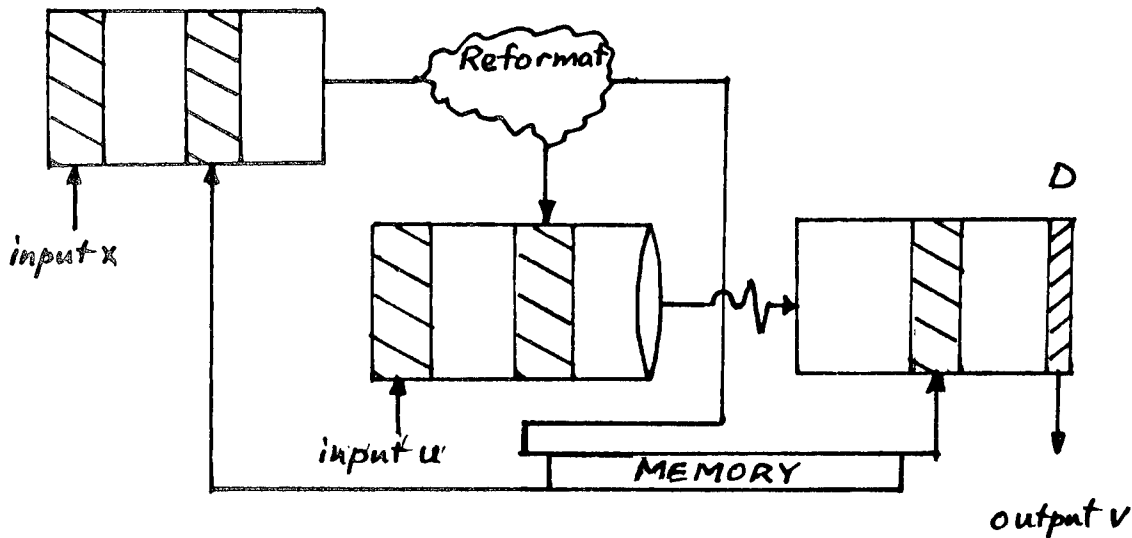
Fig. 4: An adaptation of the IP associative memory concept for sensor fusion

2 Experiment

2.1 Experimental Design

The optical associative memory was assembled as shown in Fig. 1. The 10-mW HeNe laser beam is expanded and collimated by a 1-in. beam expander. This beam traverses the two SLM's, with pixels of the first imaged onto those of the second SLM. The nature of the adding/subtracting of polarization rotations to perform a correlation was described earlier. At the Reticon array in the focal plane of the cylindrical integrating lens, all major diffraction orders of each d_k element are received by the proper detector pixels in order to represent the level of correlation approximately linearly.

S_8	0 1 1 0 1 0 0 1 0 1 0 1 1 0 1 0 0 1 0 1 0 1 1 0
V_2	0 1 1 0 1 0 0 1 0 0 0 1 1 1 1 0 0 1 0 1 0 1 1 0

1a

MEMORY VECTORS	1	2	3	4	5	6	7	8	9	10	11	12
INPUT HAMMING DISTANCES	12	12	12	12	12	12	12	2	10	10	10	12

1b

Table 1: Corruption of a memory and the Hamming distances of the corrupted memory to other memories.

In order to characterize the recognition power of the system, five experiments using varying numbers, M, of 16- and 24-bit memory vectors were conducted. Before giving the details of these experiments we note the source of memory vectors for all these experiments. These memory vectors were constructed so that the Hamming separation between any pair was exactly $N/2$, and there were equal numbers of 0's and 1's in a vector. Table 1a shows a memory vector, denoted S_8, which has a Hamming distance of $N/2 = 12$ to all other memory vectors. A version of S_8 corrupted by 2 bits is shown, denoted V_2. V_2 is presented as input to the optical associative memory

(OAM). Table 1b shows the Hamming distance between V_2 and each of the 12 memory vectors. Experiments with the OAM were paralleled by computer simulation in order to gauge the particular effects of optical imperfections.

The first two experiments were conducted to find differences in recognition caused by the position of the corrupted bits in an input. The input vectors were created by changing certain bits in a reference memory vector. In the first experiment (Exp. 1), one memory vector was corrupted one bit at a time, starting in the center of the memory vector. Four and 16 16-bit memory vectors and 6, 12, and 24 24-bit memory vectors were used. The second experiment (Exp. 2) corrupted one memory vector one bit at a time, starting from the ends of the memory vector. Four and 16 16-bit memory vectors and 6 and 12 24-bit memory vectors were used.

The third experiment (Exp. 3) was conducted to determine the response of the OAM to a situation wherein the input vector lay between two different memory vectors, with the sum of the two Hamming distances equal to $N/2$. In this experiment, the input vector was corrupted one bit at a time in such a way as to become more distant from one memory vector and at the same time approach another memory vector. This experiment was conducted with 4 and 16 16-bit memory vectors, and also with 6 and 12 24-bit memory vectors.

The fourth experiment (Exp. 4) explored the effect of varying the level of interior-plane (plane C) thresholding, with the input vectors held constant. This experiment was conducted using exclusively 12 24-bit memory vectors. The threshold was reduced to the minimum value for which the memory vector would be perfectly recognized. Data at still lower thresholds were also taken in order to observe the nature of failure due to the interference of noise, as explained earlier. This experiment was conducted with input vectors corrupted by 2, 3, or 4 bits from a particular memory vector.

The fifth experiment (Exp. 5) sought to determine the response of the OAM to 16 random vectors generated by computer. Since the threshold level was a variable in this experiment, care was taken to vary the thresholds for the OAM experiment and the computer simulation experiment in a consistent fashion. These random vectors were seen to fall into three categories as shown in Table 2. Category 1 consisted of those random vectors recognized using a low threshold value (0.55). Category 2 consisted of those random vectors requiring a higher threshold for recognition. The third category consisted of the random vectors in which the lowest Hamming distance was equal for two or more memory vectors.

Above, and in all the following, we state thresholds on a scale of 0 to 1.0. This normalization is obtained by dividing the digital value of the threshold by the digital d_p value obtained from the Reticon array (and the analog-to-digital converter) for recogni-

tion of an uncorrupted memory vector as input, i.e., $\|u_p\|^2$. Thus if perfect correlation yields 240, and the threshold is set to 120, the threshold is reported as 0.5. Values from 0 to 0.5 represent net negative correlations, while values from 0.5 to 1.0 are net positive.

	THLD 0.55	THLD 0.60	HAMMING DISTANCES WERE THE SAME FOR 2 OR MORE OF THE MEMORY VECTORS
IP	3	5	8 NO RECOGNITION
OAM	3	4 / 5	RECOGNIZED ONE OF THE LOWEST HAMMING DISTANCE VECTORS WITH THLD = 0.55 5/8
		THLD = 0.65 1 / 5	RECOGNIZED A DIFFERENT VECTOR THAN ONE WITH A LOWEST HAMMING DISTANCE WITH THLD = 0.55 2/8
			1 / 8 NO RECOGNITION

Table 2: Results of random vector response to the OAM.

2.2   Results

In the first and second experiments, the OAM usually failed to recognize an input before a Hamming distance of $N/4$ was reached between the corrupted input vector and the reference memory vector. $N/4$ is half the separation of the memory vectors. In Exp. 1, errors were introduced into a memory vector starting from the center of the vector. As the number of corrupted bits increased, the OAM consistently failed in recognition before the simulator did. Table 3a shows some results for a 24 24-bit memory vector. A memory vector denoted S_{17} was corrupted progressively one bit at a time from the center; recognition by the simulator failed when 6 bits had been changed. The criterion $N/4$ was 6 in this case. Table 3b shows results for the OAM with the same input; failure occurred when 4 bits had been changed. In this case, and in others, the simulator evidenced a higher tolerance for corruption than the OAM.

The performance of the OAM often improved as the number of memories decreased; however, this was not a strong influence, in agreement with the predicted dependence on M. In Exp. 1 the convergence length of the OAM compared to the simulator for 6, 12, and 24 memories was 80, 75, and 60 percent respectively.

The results of Exp. 2 differed from Exp. 1 mainly in that, for the OAM, a greater number of corrupted bits was tolerated before association ceased to occur. These two experiments taken together show that, in the the optical implementation where imperfect beam uniformity is a factor, the location of the corrupted bits in the SLM format is significant. Recognition usually extends further when errors are introduced from the ends than when they are introduced from the center. This was not a factor in the simulation.

In Exp. 3, in which the errors were progressively introduced in such a way as to move the input vector away from one memory and toward another, recognition was generally maintained up to $N/4$, at which point recognition of the other memory commenced. No difference between the OAM and the simulator was noticed in this respect.

COMPUTER SIMULATION

INPUT VECTOR	VECTOR RECOGNIZED
V 1	S_{17}
V 2	S_{17}
V 3	S_{17}
V 4	S_{17}
V 5	S_{17}
V 6	none
V 7	none
V 8	none
V 9	none

3 a

OPTICAL ASSOCIATIVE MEMORY

INPUT VECTOR	VECTOR RECOGNIZED
V 1	S_{17}
V 2	S_{17}
V 3	S_{17}
V 4	none
V 5	none
V 6	none
V 7	none
V 8	none
V 9	none

3 b

Table 3: Simulator and OAM response to an input (V_n, n from 1 to 9) which is a corrupted version of one particular memory vector, S_{17}.

In Exp. 4, it was observed that recognition always occurred when the threshold was made high enough; however the minimum threshold needed was not so high as predicted

by theory. The reason for this is not entirely understood. With more extensive testing we would expect to find that higher thresholds would be needed to remove almost all instances of failure. The sufficient threshold was just modestly above 0.5; this is the level needed to eliminate all d_k that represent net negative correlations. In all cases observed, the simulator recognized the corrupted vectors at a somewhat lower threshold value than did the OAM. There was a weak dependence of the threshold needed for reliable recognition on the number of corrupted bits in the input vector, at least up to $N/6$. For four bits in error, the threshold required was significantly higher than for two or three bits in error. This type of experiment was conducted only for the case of $N = 24, M = 12$.

In Exp. 5, the response of the OAM (and the simulator) to random vector inputs was seen to fall into three categories. Category 1 consisted of those random vectors recognized using a low threshold value (0.55 of the full correlation value). Category 2 consisted of those random vectors which required a higher threshold before recognition was achieved (> 0.55). Category 3 consisted of the random vectors with the same lowest Hamming separation from two or more memories. Table 2 shows the OAM and simulator results for Exp. 5. In category 1 the OAM recognized three (of 16) random vectors at a threshold of 0.55. In category 2, five vectors (of 16) were not recognized at 0.55 but required a threshold of 0.60 or 0.65. It is significant that no random vector which had a low Hamming distance from only one memory vector failed to be recognized at a moderately high threshold. Category 3 contained the 8 (of 16) vectors which by rights should not cause association. Indeed no association was made by the simulator. However, for the OAM, 5 of the 8 vectors were recognized, association being made to one of the memories which had a lowest Hamming distance. Two of the 8 led to association with some other memory. Only one vector was not recognized under any conditions. These results show that, when two or more memory vectors have the same lowest Hamming distance from the input vector, the OAM will recognize one of these vectors over others most of the time. This prejudiced association can of course be considered undesirable. It is due to variation in intensity across the optical beam, which gives biased significance to bit flips in the center of the SLM format and thus preeminence to the association with fewer central disagreements. With regard to the effect of threshold that separates categories 1 and 2, we see that a significantly higher proportion of random inputs will be associated with established memories when the threshold at plane C is available for adjustment.

3 Conclusions

We have demonstrated an optical system which implements an optical associative memory. We have shown that the ability to threshold at the intermediate plane in an inner-product architecture is an important asset with respect to recognition of corrupted vectors in a high-capacity (high-M) system. The limitations of the OAM assembled in this study lie in the difficulty of providing a uniform distribution of light and in obtaining accurate switching of the pixels of the SLM's used in this work.

References

1. W. E. Ross, D. Psaltis, and R. H. Anderson, Opt. Engin. **22**, 485–490, 1983
2. J. J. Hopfield, Proc. Natl. Acad. Sci. **79**, 2554–2558, 1982
3. J. R. Goff, paper accepted for presentation at IJCNN Conference, San Diego, June 17–21, 1990
4. D. Psaltis and C. H. Park, AIP Snowbird Conference Proc. **151**, 370–375, 1986
5. D. Psaltis and N. Farhat, Opt. Lett. **10**, 98–100, 1985
6. D. Z. Anderson, SPIE Proc. **613**, 85–92, 1986
7. B. H. Soffer, G. J. Dunning, Y. Owechko, and E. Marom, Opt. Lett. **11**, 118–120, 1986

Biodegradation of Energetic Materials (U)

*Kenneth Gold, Dr., and Bruce W. Brodman, Mr.
U.S. Army Armament Research, Development and
Engineering Center, Picatinny Arsenal, NJ 07806-5000

INTRODUCTION

The Army, in its continuing concern for environmental problems, has made a considerable investment in developing the technology and implementing production methods which minimize the potential for adverse environmental impacts. These concerns are heightened by the new legislation which has been instituted to prevent further environmental deterioration and to correct past problems[1,2,3].

The Army faces special problems in managing waste streams generated during the manufacture of energetic materials. In addition to what might be considered "normal" industrial chemical wastes, e.g., metal ions, these waste streams contain energetic contaminants as well. Such materials add a further safety risk which must be considered during storage and handling of wastes. Further, as single service manager for the production of energetic materials, the Army faces additional problems such as bioremediation of contaminated industrial sites and disposal of scrap munitions.

In general, the disposal practices in use by the Army have included incineration, burial, and chemical decomposition. These processes, however, are no longer as attractive as they once were. Open air burning may result in atmospheric pollution; burial of solids on land may contaminate streams and aquifers; chemical decomposition may result in wastes which also present environmental problems. The immediate challenge, therefore, is to find alternative non-polluting cost effective methods for degrading energetic wastes. Such methods are currently being explored by the Army. A hybrid microbiological/chemical process was used at the Badger Army Ammunition Plant (BAAP) for the decomposition of nitrocellulose (NC) fines. Holston AAP presently uses a microbiological process for the removal of cyclotrimethylenetrinitramine (RDX) from their production waste stream. The Army is also studying the use of composting

to degrade energetic materials.

The importance of exploring new technologies for the degradation of energetic wastes cannot be overstated. At present the critical military materiel trinitrotoluene (TNT) cannot even be manufactured in the United States and the explosive must now be imported. U.S. manufacturing facilities have been forced to close until new environmentally safe treatment processes are developed to cope with the volume of contaminated wastes (red water) generated during TNT production.

New advances in microbiology and in the emerging field of biotechnology have demonstrated that energetic materials are not as resistant to biological attack as previously thought. As a result, biodegradation is now seen as an attractive waste treatment alternative to earlier disposal practices.

At least three of the functional groups commonly found in energetic materials are vulnerable to biological attack -- organic nitrate esters, nitro compounds and nitramines. The focus of this research contribution is on the degradation of organic nitrate esters, in particular the dinitrate ester of diethyleneglycol [diethyleneglycol dinitrate] (DEGDN) and nitrocellulose (NC).

MATERIALS AND METHODS

1. **The Dinitrate Ester of Diethyleneglycol (DEGDN).**

Denitration experiments utilizing the enzyme glutathione S-transferase and DEGDN as the nitrate ester source were carried out in one milliliter portions of aqueous 0.1M potassium phosphate buffer (pH 7.5) containing the sodium salt of ethylenediamine tetraacetic acid (EDTA) (2mM), a chelating agent. Mixtures were incubated in a water bath at 26 $^{\circ}$C for various periods of time. In general, intervals of 15-20 minutes sufficed to demonstrate denitration effects. Definitive tests were carried out over incubation intervals of up to 40 minutes.

The dinitrate ester of diethyleneglycol was added in 5 microliter aliquots to borosilicate glass reaction vessels. Residual DEGDN was carefully removed from the test tubes at the termination of the denitration experiments just prior to making the analytical determinations. The relative amount of nitrite released from DEGDN as a result of chemical or enzymatic denitration was determined by measuring the absorbance (at 543 nm) of the red diazo dye produced by the Griess reagents in the presence of inorganic nitrite ions.

Reduced glutathione (GSH) and glutathione S-transferase were purchased from Sigma Chemical Co. The preferred concentrations of the

reactants used were: GSH,0.3mM; GSH S-transferase, 0.1mg/ml; DEGDN, 5 microliters/ml. The source of the enzyme glutathione S-transferase was rabbit liver. The enzyme preparation was used as received, and according to the accompanying description, contained 74% protein, with the balance consisting of Tris phosphate buffer salts, reduced glutathione and EDTA.

2. Nitrocellulose.

The biodegradation of nitrocellulose by the fungus Aspergillus fumigatus was carried out in 10 liters of medium in a 14 liter fermentation reactor; temperature was kept in the range of 25 to 30 °C, and the initial pH was 4.5. The culture medium, reported earlier[4], was supplemented with 3% each of glucose and double base ball propellant. A YSI glucose analyzer was used to monitor the glucose concentration at intervals throughout the incubation period.

The chemical degradation of nitrocellulose by sulfhydryl-containing compounds was performed on deterrent-free single base nitrocellulose ball propellant (13.09% N). The initial pH was in the range 7.2 to 8.2. Aqueous phosphate (0.1M) or tris(hydroxymethyl)aminomethane (0.05M) buffers supplemented with sodium EDTA (2mM) were employed[5]. The presence of nitrite in these reaction mixtures was determined spectrophotometrically at 543nm after production of a diazo dye according to standard practices.

3. Nitrite Reductase (NiR).[6]

Pseudomonas sp. was isolated from soil and the enzyme nitrite reductase induced by cultivation in nitrite-enriched medium under facultative anaerobic conditions. Twenty four to seventy two hour old cultures were used. Cells were centrifuged, washed and resuspended in cold buffer. Breakage was accomplished by means of rapid depressurization using a Parr disruption bomb pressurized with nitrogen gas at 1500 psig. The presence of NiR was confirmed by assay.

RESULTS

1. The Dinitrate Ester of Diethyleneglycol (DEGDN).

These studies demonstrate the phenomenon of enzymatic denitration of the dinitrate ester of diethyleneglycol in aqueous buffer using mammalian glutathione S-transferase. Nucleophilic attack of DEGDN by reduced glutathione is shown in Fig. 1, and the enhanced denitration effect provided by the enzyme is shown in Fig. 2. The reagent blank used for the analyses in Fig. 2 was free of DEGDN at the outset of the experiment. The "no addition" control, therefore, represents hydrolysis of the dinitrate ester in buffer.

Denitration of DEGDN by GSH
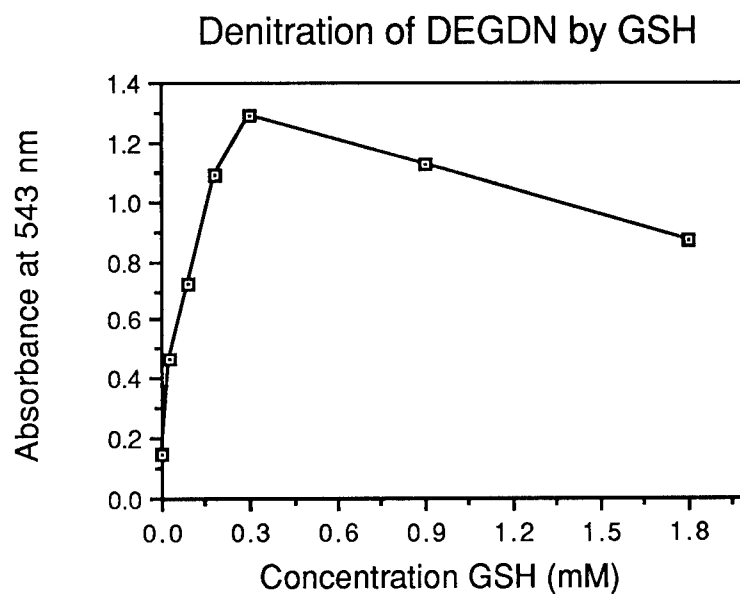
Figure 1. Denitration of DEGDN by reduced glutathione. Data points reflect the relative amounts of inorganic nitrite ions released from the ester after a 2.5 hr incubation period.

Enzymatic Denitration of DEGDN
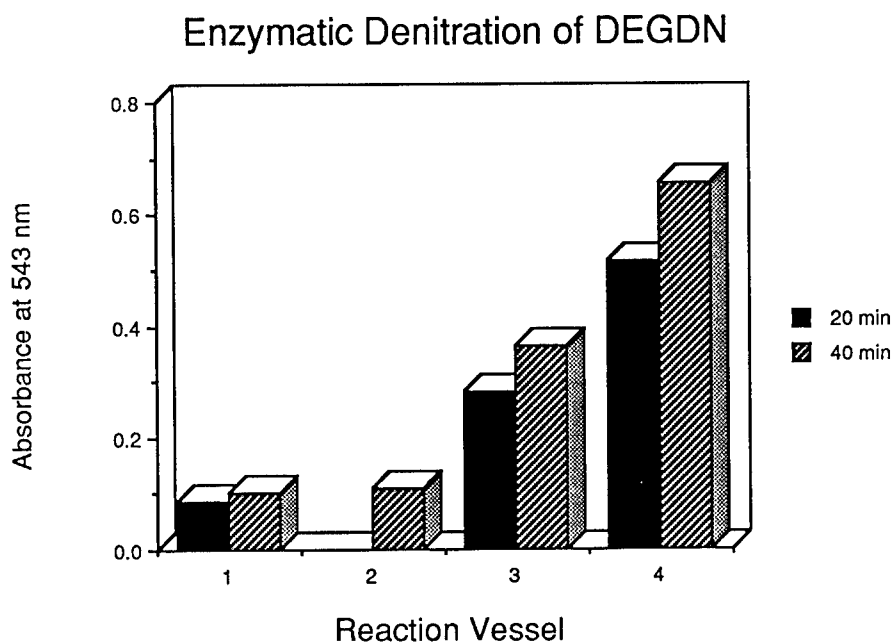
Figure 2. Enzymatic and non-enzymatic denitration of DEGDN. Assay intervals 20 and 40 minutes. DEGDN present in all reaction vessels. 1) no addition; 2) GSH S-transferase; 3) GSH; 4) GSH + GSH S-transferase.

2. Nitrocellulose

The results of a fermentation reaction carried out in cooperation with Lehigh University are given in Figure 3. The data show that the glucose concentration cycles during the course of the fermentation, the oscillations being most pronounced during the first 100 hours of growth.

Figure 4 illustrates the effect of l-cysteine on the liberation of nitrite ions from nitrocellulose an aqueous tris buffer. The denitration effect observed for l-cysteine was further enhanced by treatment at $100^{\circ}C$ and by the addition of methyl viologen (which serves as an electron carrier). The results are shown in Figure 5.

The effectiveness of l-cysteine or dithioerythritol as denitrating agents, coupled with methyl viologen as an electron carrier, is further illustrated in Table 1.

Glucose Content in A. fumigatus Culture
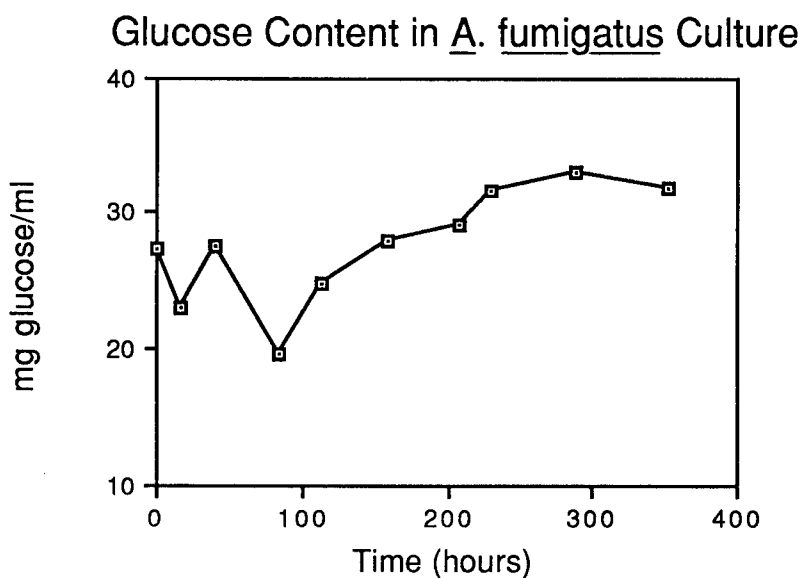
Figure 3. Analysis of glucose in culture of A. fumigatus at various time intervals after inoculating the reactor.

Effects of L-Cysteine on Denitration of NC
(Analysis for Nitrite)
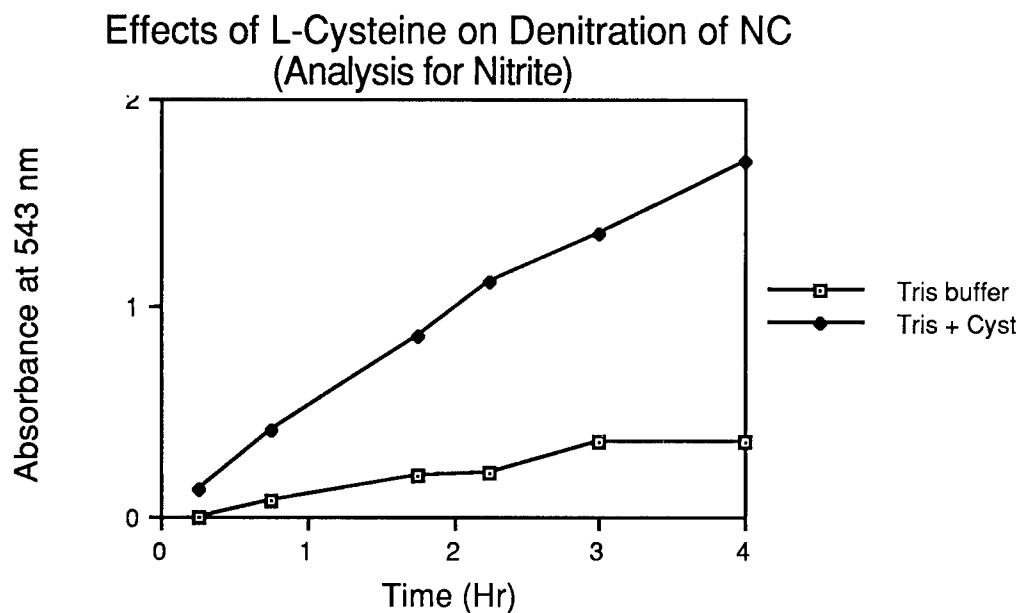
Figure 4. The release of nitrite ions from NC ball propellant due to nucleophilic attack by l-cysteine.

Effects of L-Cysteine at 100 C
(Analysis for Nitrite)
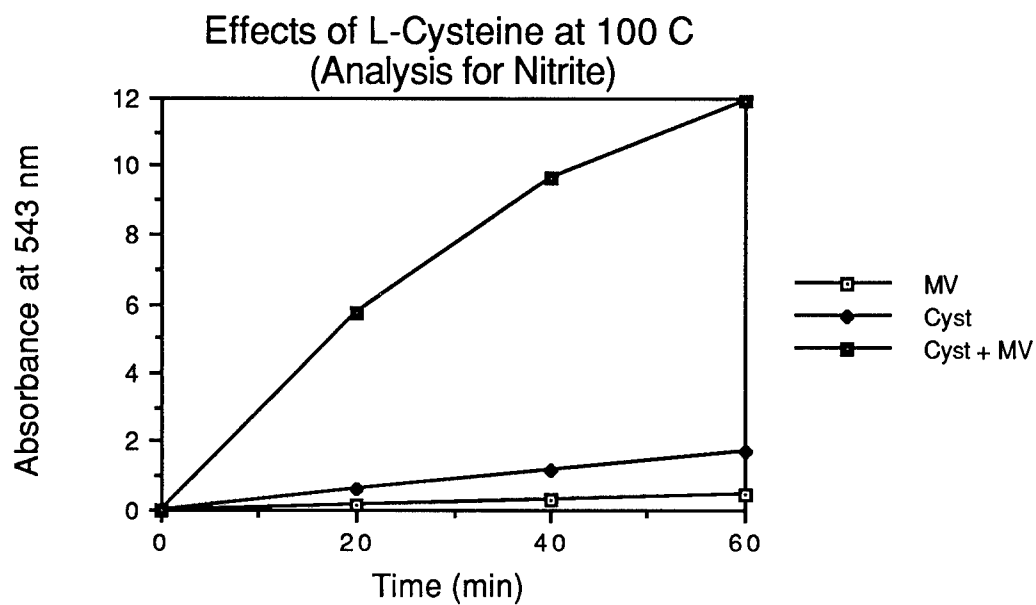
Figure 5. Denitration of ball propellant (in phosphate buffer) at elevated temperature [100 °C in the presence of l-cysteine (Cyst) and methyl viologen (MV).]

Table 1. Denitration of nitrocellulose in the presence of methyl viologen and a thiol. Conditions: phosphate buffer, 0.1M; sodium EDTA, 2mM; methyl viologen, 3mM; l-cysteine free base, 0.05M or dithioerythritol, 8.8mM. Volume of buffer, 1 liter.

	l-cysteine	dithioerythritol
Reaction period at 100 $^\circ$C	3 hr	2.5 hr
Initial pH	7.8	7.8
Final pH	7.4	7.6
Initial weight of NC	2.0 g	2.0 g
Final weight of NC	1.25 g	1.37 g
Total loss on processing	0.75 g	0.63 g
Percent lost on processing	37.5	31.5
% nitrogen in NC (initial)	13.09	13.09
% nitrogen in NC (final)	unchanged	unchanged
N in supernatant	0.102 g	0.111g

3. Nitrite Reductase

The build-up of nitrite ions in the reaction mixture normally had an inhibitory effect on the denitration reaction brought on by sulfhydryls. Such inhibition was avoided in the experiment illustrated in Table 1 through the use of a large volume of buffer solution. An alternative process was developed for future denitration reactions which are carried out at low temperatures and use sulfhydryl-containing compounds to effect the nucleophilic attack on nitrocellulose. The enzyme nitrite reductase (NiR), isolated from Pseudomonas sp., promptly and effectively lowered the nitrite content in the reaction mixture. Usually, only a few hours incubation were needed to achieve a reduction in nitrite ion concentration of 10 to 50%. The results of using nitrite reductase overnight to remove inorganic nitrite ions from a reaction mixture are illustrated in Table 2.

Table 2. The lowering of the inorganic ion concentration in a sample
of a reaction mixture as a result of incubating it overnight at 33 $^{\circ}$C in
the presence of nitrite reductase. (Control: minus NiR; Experimental,
plus NiR).

Sample #	Absorbance at 543 nm	
	Control	Experimental
1	0.86	0.55
2	0.82	0.53

DISCUSSION

Three biodegradative processes are described here which have the
potential to be utilized for treatment of wastes generated by production of
energetic materials or upon demilitarization of munitions. The first
utilizes enzymes of mammalian origin, the second intact fungal cells, and
the third an enzyme of bacterial origin in a hybrid chemical/enzymatic
process.

The release of inorganic nitrite from DEGDN as a result of the non-
enzymatic action of glutathione and the enzymatic action of GSH S-
transferase is immediately apparent in Figures 1 and 2. The reaction is
not limited to DEGDN. The reaction illustrated in Figure 2 is viewed as a
model for enzymatic denitration of other nitrate esters used in munitions.
For example, there is an extensive body of literature dealing with the
degradation of nitroglycerine (NG) by mammalian GSH S-transferase as it
applies to human physiology and the treatment of angina[7,8,9]. Needleman
and Hunter[9] proposed a denitration scheme for nitroglycerine in which
reduced glutathione was regenerated after it was oxidized in the reaction
with NG in the presence of the transferase. Additional ingredients in the
Needleman and Hunter reaction medium included glutathione reductase (GR)
and the cofactor reduced nicotinamide-adenine dinucleotide phosphate
(NADPH$_2$). These ingredients were also tested along with the DEGDN
denitration experiments described above, but further work with this system
is needed to determine optimum assay conditions for their use.

Other recent contributions relevant to biodegradation of nitrate esters
worthy of mention are those of Ropenga and Lenfant[10,11] who modified

isosorbide dinitrate to the mononitrate ester by incubating it with various bacterial and fungal cultures as well as the GSH S-transferase positive protozoan Tetrahymena thermophila. Langston-Unkefer[12] recently isolated microbial species capable of degrading NG; however, the metabolic pathways have not yet been identified.

Until recently, there have been relatively few studies dealing specifically with the total biodegradation of nitrocellulose. The study by Brodman and Devine[4] on the denitration of nitrocellulose stands out as one of the earliest to demonstrate that the polymer can be successfully denitrated. Brodman and Devine first hypothesized that denitrated nitrocellulose would be an effective substitute for nitrocellulose impregnated with a chemical deterrent as a means of controlling burning rate in small arms propellants. To test this hypothesis, a selective denitration process was chosen which employed the fungus Aspergillus fumigatus. They concluded from the results that the denitration mechanism involved aqueous hydrolysis of the nitrate ester group followed by utilization of the nitrogen. They postulated that the carbon backbone remained intact. Il'inskaya and Leshchinskaya[13] have confirmed Brodman and Devine's findings for Aspergillus and they commented on the usefulness of microorganisms for treatment of NC production wastes. They also reported the presence of an enzyme pathway in the bacterium Pseudomonas fluorescens capable of denitrating nitrocellulose.

The experimental work with A. fumigatus was continued in cooperation with Lehigh University. The data reported here (Figure 3) show that glucose concentration oscillates in the culture medium. The cycle is most pronounced within the first 100 hours after inoculation. This cycling apparently is the result of glucose utilization by the fungus resulting in further glucose hydrolysis from the denitrated portion of the cellulose backbone. This process, therefore, was a total biodegradation of NC and perhaps could be optimized to go totally to gaseous products.

The hybrid chemical/enzymatic process described here was developed as a method for denitrating nitrocellulose wastes (Figures 4-5)[5]. The effectiveness of the chemical denitration reaction employing organic sulfhydryl-containing compounds such as the amino acid l-cysteine (the dithiol dithioerythritol is also effective) coupled with methyl viologen is illustrated by the data in Table 1. There was a substantial loss of ball propellant mass, amounting to approximately one third of the initial charge in the reaction vessels. Although inorganic nitrite was released during the reaction period, the percentage nitrogen in the nitrocellulose was unchanged at the end of the experiment. This is attributed to the combined effects of denitration resulting from nucleophilic attack by the thiol followed by removal of glucose molecules from the cellulosic backbone by hydrolysis. (The oxidation product in the l-cysteine reaction is l-cystine which is poorly soluble in the reaction mixture. Hence, this insoluble

amino acid can easily be removed for recycling by filtration.)

The use of the enzyme nitrite reductase in conjunction with this NC degradative technique has some advantages over earlier degradative methods. Table 2 demonstrates the feasibility of using the extract of Pseudomonas sp. to achieve a sizeable reduction in nitrite ion concentration in a thiol-containing reaction mixture (dithiothreitol). This process in combination with the chemical degradation utilizing sulfhydryl-containing compounds would be relatively mild, carried out in aqueous potassium phosphate buffer at pH 7.2 to 8.2 at temperatures in the range 25 to 37 $^{\circ}$C. Moreover, the use of specific enzymes such as nitrite reductase has advantages over utilizing microbial biomass to accomplish a given digestion reaction. Enzymes provide a high degree of reproducibility and control over the reaction conditions; they are specific and recoverable; they can be immobilized for convenience. Bacterial and fungal assemblages, on the other hand, may vary greatly in composition depending upon a number of factors such as nutrient content, inhibitory substances in the reaction vessel, and temperature fluctuations.

The enzymatic reduction of inorganic nitrite potentially has broad applications in the energetic materials field. Nitrite ions are contaminants present in red water, as well as in other waste streams resulting from chemical degradation of energetic materials. Treatment of waste streams with this enzyme is seen as a way to decrease toxicity of the mixture to microbial species used in on-stream biodegradation processes. For example, the fungus, Phanerochaete chrysosporium (white rot fungus) which degrades pink water, is one of the primary candidates under consideration to biodegrade red water. However, nitrite ions at the concentration found in red water may be toxic to this fungus. Adjustment of the nitrite ion concentration using the nitrite reductase enzyme process should remove this potential source of toxicity, and thus enable the fungus to attack and digest the organic components of the red water.

Biotechnology is in its infancy, but the preliminary data indicate that it will have important applications in the energetic materials field. The new chemical and biodegradative processes described here have been proposed to the U.S. Army for bioremediation of contaminated soil at the Badger Army Ammunition Plant, and also for bioremediation of the burning grounds at Picatinny Arsenal. They are also viable alternatives to open air burning of energetic materials, and, therefore, are expected to contribute substantially to improving the environment at propellant and explosive manufacturing facilities. A further application of this technology involves the specific molecular transformation of nitrate ester compounds in order to enhance their effects as long lasting cardiovascular dilators useful in the treatment of angina[10,11].

CONCLUSIONS

1) The mammalian enzyme glutathione S-transferase was used to denitrate the dinitrate ester of diethyleneglycol (DEGDN).

2) The fungus _Aspergillus fumigatus_ degraded nitrocellulose to glucose with some subsequent glucose utilization.

3) Organic sulfhydryl-containing compounds were used for the total degradation of nitrocellulose.

4) The enzyme nitrite reductase effectively lowered the nitrite ion concentration in reaction mixtures used to degrade nitrocellulose.

REFERENCES

1. The Resource Conservation and Recovery Act. Public Law 94-580, 94th Congress. 1976.

2. Toxic Substances Control Act. In House Committee on Energy and Commerce. Compilation of Selected Acts within the Jurisdiction of the Committee on Energy and Commerce. Vol. 3 Environmental Law. Committee Print, 97-C. 97th Congress, 1st Session. 1981.

3. Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (Superfund). In House Committee on Energy and Commerce. Compilation of Selected Acts within the Jurisdiction of the Committee on Energy and Commerce. Vol. 3 Environmental Law. Committee Print, 97-C. 97th Congress, 1st Session. 1981.

4. Brodman, B. W. and Devine, M. P., "Microbial Attack of Nitrocellulose," Journal of Applied Polymer Science, Vol. 26, pp. 997-1000. 1981.

5. Gold, K. and Brodman, B. W., "Chemical Process for the Denitration of Nitrocellulose," U. S. Patent, Number 4,814,439. March 21, 1989.

6. Gold, K. and Brodman, B. W., "Enzymatic Process," U. S. Patent, Number 4,756,832. July 12, 1988.

7. Oberst, F. W. and Snyder, F. H., "Studies on Nitrate Esters 1. Nitrite-Producing Systems in Rabbit Tissues," Journal of Pharmacology and Experimental Therapy, Vol. 93, pp 444-450. 1948.

8. Hepple, L. A. and Hilmoe, R. J., "Metabolism of Inorganic Nitrite and Nitrate Esters II. the Enzymatic Reduction of Nitroglycerin and Erythritol Tetranitrate by Glutathione," Journal of Biological Chemistry, Vol. 183, pp. 129-138. 1950.

9. Needleman, P. and Hunter, F. E., Jr., "The Transformation of Glyceryl Trinitrate and Other Nitrates by Glutathione-Organic Nitrate Reductase," Molecular Pharmacology, Vol. 1. pp. 77-86. 1965.

10. Ropenga, J. S. and Lenfant, M., "Bioconversion of Isosorbide Dinitrate into Isosorbide Mononitrate by the Protozoan Tetrahymena thermophila: Relationship to Glutathione Transferase Levels," Applied Microbiology and Biotechnology, Vol. 26, pp. 117-119. 1987.

11. Ropenga, J. S. and Lenfant, M., "Bioconversion of Isosorbide Dinitrate by Various Microorganisms," Applied Microbiology and Biotechnology, Vol. 27, pp. 358-361. 1988.

12. Langston-Unkefer, P. J. Unpublished data.

13. Il'inskaya, O. N. and Leshchinskaya, I. B., "Action of Microorganisms on Cellulose Nitroesters," Biotechnologiya, Vol. 4, pp. 495-500. 1988.

Soldier Performance Research Project:
Armor Field and SIMNET Tests (U)

*Scott E. Graham, Dr.
U.S. Army Research Institute
Fort Knox, KY 40121-5620

The Armor force with its M1 Abrams tank continues to grow more sophisticated and lethal. Despite the technological advances in hardware, the effectiveness of the weapon systems is directly determined by the skills and resourcefulness of the Armor crewmen. Furthermore, the success of Armor tactical operations requires mentally alert soldiers who can seize and hold the initiative on an increasingly complex battlefield. If the Army is to maximize the effectiveness of the Armor force, with its $2.5 million tank, the Army must maximize the skills of its Armor crewmen.

The research reported here was conducted as part of a effort directed by the Commanding General, Training and Doctrine Command (TRADOC) to ensure that the best soldiers are operating the high-tech weapon systems in the Army inventory. The goal of the Soldier Performance Research Project (SPRP) is to conduct more rigorous tests and analyses of the cognitive skill requirements of first and second term soldiers. SPRP results are being used to determine the Army's aggregate recruiting quality needs and the TRADOC distribution of quality requirements.

The crux of the SPRP was to determine whether soldiers with high mental abilities do, in fact, perform better on combat tasks than soldiers with lower mental abilities. The primary drawback to increased manpower quality is that soldiers with higher mental abilities cost more to recruit than do soldiers with lower mental abilities. The critical question is whether it is cost-effective to recruit smarter, albeit more costly, soldiers. If there is little difference in performance as a function of mental ability, the Army would be better off spending its limited funds on other programs. On the other hand, if combat performance and weapon system effectiveness are significantly enhanced by high ability soldiers, recruiting quality soldiers could be the most cost-effective strategy for increasing system performance.

The issue is complicated by the fact that soldiers, in the present case 19K M1 tank crewmen, are required to perform hundreds of tasks under a variety of condi-

tions. Clearly, mental ability is not going to be related to performance of all tasks all of the time. To cover as many Armor tasks as possible, the Armor portion of the SPRP included two phases. The two phases sampled various aspects of the armor crewman task domain and tested soldiers at different points in their enlisted careers. Phase I examined the impact of soldier quality on the gunnery performance of soldiers enrolled in Armor One Station Unit Training[1]. Phase II, reported here, tested the collective combat skills of reconstituted tank crews using first and second term soldiers from U.S. Armor units. Two separate tests were administered including a high combat realism Field exercise and a platoon tactical exercise in the Simulation Networking (SIMNET) system.

Soldier Quality and Armor Performance

All soldiers enlisting in the Armed Services are given a standardized paper-and-pencil test, the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB contains ten cognitive subtests which are combined to form a number of composites. One composite, the Armed Forces Qualification Test (AFQT), is used by the Army to classify soldiers into mental categories. The specific categories used in the SPRP are I & II combined, IIIA, IIIB, and IV. As a note, the Army cannot, by law, enlist category V's nor more than 20% category IV's in an accession year.

A number of research efforts in the past decade has examined the relationship between soldier quality and Armor performance, with the majority of the research dealing with Armor gunnery performance. For the most part, soldier quality has consistently correlated with the performance of tasks other than live-fire gunnery. Several research efforts have, however, found relationships between soldier quality and non-gunnery Armor performance. Tziner and Eden[2] manipulated the composition of three-man Israeli tank crews by high and low mental ability and motivation. The results showed an additive effect of mental ability, i.e., the more high ability soldiers in the crew, the better the performance. Perhaps more interesting was the interaction among the ability levels of the three crew positions. Crews composed of three high ability soldiers were ranked more effective than expected, while crews composed of three low ability soldiers were ranked less effective than expected.

Other research has looked at individual soldier quality and armor performance. Black and Mitchell[3] found a strong relationship between gunner AFQT scores and performance on an M1 computer panel test. Graham[4] found soldier quality to be related to the command, control, and communication (C^3) performance of M1 TCs during single tank tactical exercises on the Simulation and Combined Arms Trainer (SIMCAT). Significant differences in performance as a function of mental ability were found on numerous task measures including: accuracy of combat reports, number and accuracy of fire commands, decoding and plotting minefield coordinates,

and the calling and adjusting of indirect fire. Regression analyses showed that the effects of mental ability greatly outweighed the effects of experience.

Purpose of Armor SPRP: Phase II

The Phase I Armor SPRP demonstrated the effects of soldier quality on the individual gunnery performance of soldiers enrolled in initial-entry training. Phase II was designed to examine the effects of soldier quality on the collective performance of first- and second-term Armor crewman with the emphasis on C^3 performance. Specifically, the purpose of the research was to evaluate the effects of AFQT category on the collective performance of armor crews in a high combat realism single tank tactical field exercise and in a SIMNET platoon tactical exercise.

METHOD

Participants

The SPRP tested 120 19K Tank Commanders and 120 19K drivers from five divisions in the continental U.S. Approximately 130 support personnel were required to construct and execute the Armor SPRP Field and SIMNET tests. The tests were administered by the Test and Experimentation Command (TEXCOM) Armor and Engineer Board.

Design

The SPRP Field and SIMNET tests were designed to assess the impact of tank crewman mental category on the collective performance of the four-man M1 tank crew. Four (AFQT) category groups were used: I & II combined, IIIA, IIIB, and IV. Two crew positions were examined in the current research, the TC and driver. Test surrogates were used in the other two crew positions as gunners and loaders. The loader and gunner surrogates simplified the design by reducing the number of mental category combinations, i.e. 4^2 rather than 4^4. The TC and driver were selected as the tested crew positions because they represent first- and second-term soldiers.

The primary experimental design for the Field and SIMNET Tests was a randomized 4 X 4 factorial design with between-subjects factors of TC AFQT category (I&II, IIIA, IIIB, and IV) and driver AFQT category (I&II, IIIA, IIIB, and IV). The level of measurement was the performance of the tank crew.

Selection rules were established to ensure the counterbalanced assignment of soldiers. First, an attempt was made to select an equal number of TCs and drivers

from each of the mental categories from each division. Second, TCs were selected such that the ranks of SSG and SGT were maintained at a 2:1 ratio. Third, TCs and drivers were paired in such a way as to systematically fill the 16-cell (4 X 4) design. The result was that no one unit provided a disproportionate number of crews to any cell. This procedure was adopted as an attempt to counterbalance the effects of unit training. Fourth, for logistical reasons, TCs were paired with drivers from the same division, with the restriction that they could not be from the same tank crew.

FIELD TEST

Description

The SPRP Field test consisted of a high combat realism single tank tactical exercise which evaluated the speed and accuracy of the tank crew performance. The scenario was developed in the context of the third day of the war, when reconstitution of tank crews would be necessary. The course extended approximately 15 kilometers primarily within the Wilcox Range area of Fort Knox, KY. While the course and its events appeared continuous to the tested crew, it was constructed as a series of stations. A brief description of each station is presented below.

Station 1. Brigade Support Area (BSA)- The first station replicated, as nearly as possible, a BSA deep within a combat environment. The tactical station covered several acres and included organizational maintenance assets, refueling trucks, a small arms repair tent, and was protected by concertina wire and armed guards. At the BSA the TC was given an operations order and told to prepare his M1 tank for combat. The crew was required to conduct preventive maintenance checks and services (PMCS), upload vehicle stores, refuel, boresight the main gun, and enter data into the ballistic computer.

Station 2. Surprise Engagement with Disabled T-72 and T-72 in Overwatch - At a designated point on the course road, the loader (a test surrogate) identified two mock Soviet tanks at approximately 1600 meters, one disabled and one in overwatch. The tank crew was required to engage the most dangerous target first, execute acceptable combat driving, and send a spot report on the radio which included the grid coordinates of the enemy vehicles. Throughout the test, the tested tank and OPFOR vehicles fired blank Hoffman charges. The opposing force (OPFOR) vehicles simulated being hit by using smoke grenades.

Station 3. Antitank Guided Missile (ATGM) Ambush in a Minefield - As the tank entered a cleared lane in a friendly minefield, it was engaged by an ATGM from a partially concealed mock soviet infantry fighting vehicle, a BMP. The station

was designed to increase cognitive workload by requiring the crew to attend to the minefield and the ATGM attack simultaneously. The TC was required to direct the driver through the minefield, issue a fire command, and continue to evaluate the situation until the engagement was resolved.

Station 4. Meeting Engagement with Enemy Stragglers - Loader Killed - At this station, the tank crew acquired three enemy soldiers at approximately 40 meters who engaged the tank with automatic rifle fire. The crew had to direct machine gun fire on the enemy soldiers. During the engagement the loader was killed, which was simulated by squirting arterial blood from a tube attached to the surrogate's neck. The crew then had to evacuate the loader's body, reconfigure as a three-man crew, and issue a proper casualty report to higher headquarters.

Station 5. Military Police Traffic Control Point - As the tank approached the TCP, the crew had to recognize the military police as friendly, issue a proper challenge, and identify the tank's correct location on the map.

Station 6. Meeting Engagement with T-72 Tank and BMP Infantry Fighting Vehicle at Short Range - The TC had to engage multiple targets from a three-man crew configuration. The TC had to recognize that the T-72 had not been killed, and re-engage. As with each of the engagements, the TC had to send a spot report including the grid coordinates of the destroyed targets.

Station 7. Automatic Weapons Ambush - TC and Gunner Killed - The tank was ambushed with automatic weapons fire by an enemy infantry squad. The gunner (test surrogate) pulled the TC's communications cord and held a card up which indicated he had been killed. The driver then became the only surviving crew member and, on his own initiative, had to move the tank out of the kill zone, determine crew status, and submit a report indicating casualties and the vehicle location.

Station 8. End of Course - As the driver proceeded to the release point, he was stopped by test controllers and asked to show his location on the map. The crew was then taken to a debriefing tent where each of the stations was discussed.

Field test scoring procedures

The Field test events were patterned after the Tank Tactical Tables in FM 17-12-1. The task lists for each station and the scoring criteria checklists were approved by the Weapons Department and the Command and Staff Department, U.S. Army Armor School. The Field Test checklist contained 125 Yes/No items which corresponded to the tasks at each station. The 125 Yes/No items were the categorized into 18 task clusters by analysts from the Armor School. Table 1 lists

the 18 task clusters along with a brief description of each. The primary performance measure for the field test was Field Test Total score which was the mean of the 18 task clusters, i.e., each of the clusters was equally weighted. Three task composites were also calculated: Precombat, Command and Control, and Communications.

Multiple data collectors located across the test course recorded performance data on feeder checklists which were later compiled onto the scoring checklist. The majority of the data was recorded by data collectors located in vans located at the top of Wilcox Range at Observation Post (OP) Alpha. Much of the data resulted from the monitoring of the tank's intercom. This was accomplished by hooking up a jump radio to the intercom system. The jump radio allowed OP Alpha to score most of the command and control tasks, e.g., issuing fire commands and directing tank through minefield. OP Alpha personnel also served as the unit to which the crew sent all reports, thereby enabling them to score the radio reports.

Table 1. Field Test Task Clusters and Descriptions

Task Clusters	Description
1. Uses TM (-10) for PMCS	Uses TMs for preoperational checks
2. Preparation of Weapon Stations	Prepares, inspects, and tests weapons
3. Troubleshooting	Identifies/corrects induced malfunctions
4. Enter Ballistic Computer Data	Conducts computer self-test
5. Boresighting Main Gun	Boresights main gun to Armor standard
6. Vehicle Load	Ensures correct load of ammo, fuel, etc.
7. Issuing Proper Fire Commands	Gives complete fire command
8. Target Engagement Procedures	Uses proper engagement procedures, e.g., engaging most dangerous target first
9. Decoding/Plotting Map Coordinate	Correctly decodes and plots minefields
10. Directing Tank through Minefield	Directs driver through minefield
11. Submits Spot Report without Cue	Submits report after each engagement
12. Accuracy of Spotrep	Issues doctrinally accurate spot reports
13. Submits Casualty Report	Submits casualty report without cue
14. Accuracy of Casualty Report	Issues accurate casualty reports
15. Troop Leading Procedures	Briefs crew on mission, conducts drills
16. Security	Issues proper challenges and passwords
17. Position Location	Reports accurate grid coordinates
18. Combat Driving	Provides stable platform, proper evasion

SIMNET TEST

The SIMNET test was designed to test the performance of tank crews within the context of platoon tactical gunnery engagements. The SIMNET system was used in lieu of a separate platoon-level field test because of cost constraints.

The SIMNET system at Fort Knox contains a local area network of combat simulators which was largely designed as a part-task tactical trainer for armor and mechanized infantry units. Each M1 tank simulator consists of a separate module with space for the four tank crewmen. The crew views computer-generated images in the tank sights and vision blocks, as well as senses computer-generated sounds and vibrations. SIMNET can support free play force-on-force tactical scenarios, in such a way that opposing sides each see their own side as friendly, e.g., M1s, and the opposite side as an opposing force, e.g., T-72s.

The SIMNET modules are equipped with intercoms and radios to support communications within and between simulators. Weapon systems and their effects are simulated such that simulators can kill or be killed. Logistic and maintenance functions are also represented such that a simulator can break down or run out of fuel or ammunition. Fire support is controlled from microcomputer stations collocated with a simulated Tactical Operations Center. Data collection capabilities include a Data Logger which maintains a computer record of activities that occur during a SIMNET exercise and a Plan View Display which provides a graphic map display of activities. For a more complete description of the SIMNET capabilities and potential training applications, refer to the SIMNET Users' Guide[5].

Test Description

The SIMNET test was similar to the Field test in a number of respects, primarily in that it assessed the speed and accuracy of a tank crew to move, shoot, and communicate. The tested crew in the SIMNET test served as a wingman to the platoon sergeant in a platoon tactical exercise, rather than as a single tank in the Field test's single tank tactical exercise. Like the Field test, the SIMNET test employed surrogate gunners and loaders. The SIMNET surrogates were different soldiers than those in the Field test. The SIMNET test took approximately one hour to run. Because only a few of the soldiers had prior SIMNET experience, the soldiers were given considerable SIMNET training prior to the testing. Following the training, the crews were tested on a 30 minute certification course in which they were required to demonstrate proficiency in the areas of navigation, vehicle identification, and use of the SIMNET communication systems. Those crews who failed to meet the criteria were given additional training and retested until they qualified.

Event 1. Prior to the test, the crew was given an operations order. The tested tank served as a wingman to the platoon sergeant. Event 1 consisted of a tactical road march in which the crew executed various platoon formations and action drills.

Event 2. The platoon formation crossed the line of departure. The TC was asked to send a spot report which included the bride coordinates.

Event 3. The platoon had a meeting engagement with an enemy tank platoon. The tested tank had to give a contact report, conduct a movement by bounds, engage the enemy tank platoon, and send a spot report.

Event 4. The platoon was attacked by an ATGM from a BMP. During the attack the platoon sergeant's tank was destroyed. The tested tank was required to issue a contact report, conduct a contact drill, engage, and report. Platoon movement resumed with the tested tank then serving as the platoon sergeant.

Event 5. The crew had to react to an ATGM ambush by issuing a contact report and fire command. As per unit standard operating procedure, the platoon sergeant (now the tested tank) had to send all reports to higher headquarters.

Event 6. The crew had to react to indirect fire and report.

Event 7. The crew assumed a hasty battle position and engaged a reinforced - rifle company as part of the platoon. The platoon leader issued a depth fire command to which the tank had to engage the last tank first and work forward. The platoon then engaged a second offensive formation, during which the platoon leader's tank was destroyed and the wingman took a mobility kill. The tested crew had to consolidate, reorganize, and report.

Event 8. The company commander instructed the tank to call indirect fire on any future targets they might encounter. Upon acquisition of four BMPs, the crew had to contact the Fire Support Team (FIST) and call and adjust indirect fire. After firing for effect on the BMPs, the tested tank observed a missile coming straight its driver's vision blocks and was destroyed.

Event 9. The TC was taken from the M1 module and tested on radio authentication challenges and the ability to correctly encode and decode grid coordinates.

A SIMNET test scoring checklist was developed similar to the field test. The checklist contained 123 Yes/No items which were categorized into 11 task clusters. Like the field test, the primary performance measure for the SIMNET test was SIMNET Total Test score which was the mean of the 11 task clusters.

The SIMNET data were gathered from several sources. As with the field test, the majority of data came from the monitoring of the tank intercom and from the radio transmissions to higher headquarters. A data collector, located at the SIMNET test Tactical Operations Center (TOC), recorded responses based on crew communications. Also, located at the TOC was the Plan View Display (PVD) which displayed a birds-eye-view of the battlefield, including the location of all elements, firer-target pairings, and the orientation of hulls and turrets. A third control device at the TOC called a shadow box contained four visual displays which paralleled the tested driver's three vision blocks and the tested TC's center vision block. Following the runs, the SIMNET Datalogger was used to analyze certain aspects of the crew's performance. For example, a plot was made of the platoons movement which showed whether the tested tank maintained the proper position in the various formations. Because the pace of the SIMNET test was largely determined by the speed of the platoon leader, a test surrogate, a total SIMNET time was not collected.

RESULTS

Field Test

Total Score. Performance on the SPRP field test was highly related to the mental categories of the crewmen. Table 2 shows the mean field test Total Score for the TC and driver mental category groups. The mean Total Score for the TC groups is also shown relative to the performance of mental category I & II TCs. Recall that the field test Total Score resulted from an equal weighting of the 18 task clusters. An Analysis of Variance (ANOVA) revealed that the main effect for TCs was significant, $F(3,116) = 5.27$, $p < .002$, while the main effect for drivers approached significance $F(3,104) = 2.35$, $p < .08$. A Newman-Keuls post hoc test indicated that Cat I&II and IIIA TCs were more accurate than Cat IIIB and IVs ($p < .05$).

Table 2 shows that Cat I&II and IIIA TCs performed virtually at the same level and that Cat IV TCs performed at only 81% of the level of Cat I&II TCs. The performance of crews as a function of the driver's mental ability is somewhat less straightforward. Nevertheless, the crews with Cat I&II drivers had the highest scores while the crews with Cat IV drivers had the lowest scores. As a reminder, the experimental design systematically paired the TCs and drivers by mental categories. This means, for example, that the mean score for Cat I&II drivers is from crews with nearly equal numbers of Cat I&II, IIIA, IIIB, and IV TCs.

As reflected in table 2, the effects of TC mental ability had a larger influence on field test performance than did driver mental ability. This is to be expected since the TC is primarily responsible for the performance of the tank. In addition, the

majority of the field test tasks directly assessed what the TC said and did. The effects of the TC and driver mental ability were, however, found to be additive. The higher the mental category of either crewman, the better the performance tended to be. The ANOVA substantiated the additive effect by finding a zero interaction between TC and driver mental ability, $F(9,104)=.30$, $p<.98$. No significant interactions of TC and driver mental ability were found in any of the SPRP analyses.

Table 2. Field Test Total Scores by Mental Category Groups

	Mental Category			
	I&II	IIIA	IIIB	IV
TC	.53 (n=32)	.52 (n=27)	.46 (n=32)	.43 (n=29)
Driver	.52 (n=32)	.47 (n=28)	.50 (n=30)	.45 (n=30)
Relative to Cat I&II TCs	100%	98%	87%	81%

Task Cluster Composites. The same pattern of mental category effects was found for the three field test task cluster composites. Of particular interest are the results of the Precombat Composite. An ANOVA showed that performance varied as a function of the TC mental category, $F(3,104) = 3.07$, $p<.03$, with the means for TC mental category I & II, IIIA, IIIB, and IV being .53, .48, .46, and .37, respectively. The Precombat composite assessed the crew's ability to prepare the M1 tank for combat. The composite contained a number of tasks which generally would be thought to correlate with mental ability, e.g., relatively complex procedural tasks, troubleshooting, and working with the ballistic computers. The results show that Cat IV TCs performed at 70% the level of Cat I&II TCs on this highly critical set of tasks. For boresighting alone, Cat IVs correctly boresighted to the Armor standard (+/- .3 mils) only 45% as often as Cat I&IIs.

Performance times. Performance speed was collected in addition to accuracy. In the operations order, the crews were instructed to engage all targets, but to get to the release point as quickly as possible. The Total Time measure did not include the 90 minutes in the BSA, but was measured from when the first target was acquired until the end of course. The mean Total Time for crews with Category I & II, IIIA, IIIB, and IV TCs was 36.6, 38.8, 40.0, 45.3 minutes, respectively. Interestingly, Total Time was significant for TC mental category, $F(3,90) = 3.11$, $p<.03$, but not for driver mental category, $F(3,90) = 0.57$, $p<.71$. The findings suggest that Total Time was more of a measure of C^3 factors rather than how fast the driver drove.

Speed/Accuracy Composite. A more comprehensive measure of combat effectiveness than either speed or accuracy alone combines the two together. One approach is to calculate the number of task clusters correctly performed per minute, i.e., Total Score divided by Total Time. The figure below shows the field test Speed/Accuracy Composite for the TC mental category groups relative to I & IIs. An ANOVA revealed significant differences for TCs, $F(3,90) = 5.02$, $p<.003$, but not for drivers, $F(3,90) = 0.83$, $p<.48$. The speed/accuracy results even more dramatically show that crews with Cat I&II and Cat IIIA TCs perform equivalently with a sharp drop in the performance of crews with Cat IV TCs.
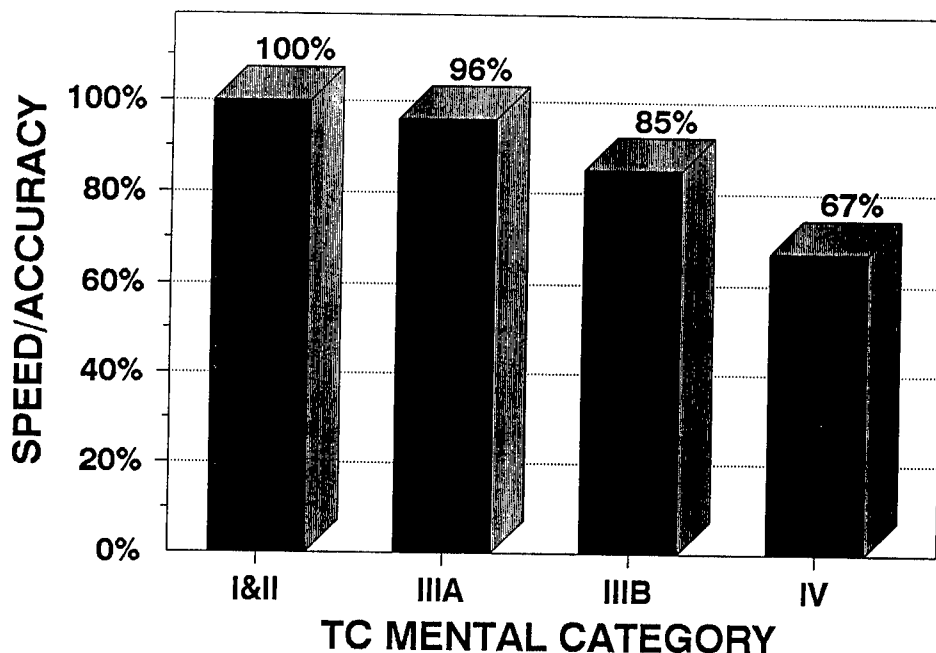
Figure 1. Field test Speed/Accuracy Composite for TC mental category groups.

Grid deviation errors. During the SPRP field test, there were five instances in which TCs were required to give grid coordinates. Four of the cases required the

TC to identify the location of enemy targets as part of spot reports. The fifth instance was when the MP at the traffic control point asked the TC to give the grid of their current location. The mean grid deviation errors in meters for the four mental category groups of TC were 736, 1004, 830, 1339. The data show that Cat IV TCs made considerably greater grid deviation errors than the other TCs. Furthermore, the overall magnitude of the errors, nearly one kilometer on average, suggests tankers have trouble determining accurate grid coordinates.

SIMNET Test

The results of the SIMNET test generally reflected the same pattern as the field test. The magnitude of differences between mental category groups were not, however, as large as found in the field test. Table 3 shows the mean SIMNET test Total Scores for TCs and drivers by mental category. Like the field test, an ANOVA found significant differences for the TCs, $F(3,104) = 3.72$, $p < .02$, and the main effect for driver mental category approached significance, $F(3,104) = 2.32$, $p < .08$. A Newman-Keuls post-hoc test found the Cat IV TCs to be lower than other three groups. The additive effects of crew mental ability again were indicated by a lack of an interaction between TC and driver mental ability, $F(9,104) = .77$, $p < .64$.

Table 3. SIMNET Test Total Scores by Mental Category Groups

	Mental Category			
	I&II	IIIA	IIIB	IV
TC	.61 (n=32)	.59 (n=27)	.58 (n=32)	.53 (n=29)
Driver	.62 (n=32)	.56 (n=28)	.57 (n=30)	.56 (n=30)
Relative to Cat I&II TCs	100%	97%	95%	87%

The SIMNET results revealed several instances where there were large performance differences as a function of mental category. The Call For Fire Accuracy task

assessed the effectiveness of the call for fire procedures, i.e., whether the target was destroyed by the adjusted artillery fire. Category IV TCs performed at only 67% of Category I&II TCs. There was also a large difference in the accuracy of encoding and decoding grid coordinates, with Cat IV TCs performing at 42% of Cat I&II TCs.

<u>Regression Analyses</u>

A series of multiple regression analyses were also conducted to estimate the separate and collective contributions of other variables in predicting SPRP performance. The criterion variable used was the Field/SIMNET Combined score. For the first set of regressions, five sets of predictors were used: TC and driver AFQT, TC and driver rank, TC and driver high school graduation status, TC and driver time in service (TIS), and unit membership. The only significant predictors of the Field/-SIMNET Combined score were the AFQT scores of the TC and driver, the rank of the TC, and the unit to which the crew belonged. Crews from the best performing unit had just returned from the National Training Center prior to the SPRP testing. After controlling for the effects of all of the other variables, the mental ability of the TC accounted for 13.5% of the variance on the combined Field and SIMNET test, while the mental ability of the driver accounted for 4% of the variance. Separate regressions also showed that mental ability of the TC and driver together accounted for 17.5% of the variance and that unit membership (a measure of training) uniquely accounted for 18% of the variance.

DISCUSSION

The Armor SPRP test results consistently demonstrate that tank crew effectiveness is influenced by the quality of the crewman. Differences in performance as a function of mental ability were not only found for the overall performance measures, but for the precombat, C^2, communications, call for fire, grid coordinate determination, and encoding/decoding tasks as well. Analyses of the field test speed/accuracy composite showed that crews with Cat IV TCs performed at 67% of crews with Cat I&II TCs. The results also clearly show that soldier quality affects the collective performance of the crew, not just the performance of individual tasks. The regression analyses showed that the mental categories of both the TC and driver were related to crew performance, with TC and driver AFQT scores accounting for 17.5% of the test variance. The effects of mental ability for the TC and driver were also found to be additive, i.e., the more smart crewmen in a tank, the better the performance of the tank. The mental category of the driver was found to influence crew effectiveness even though the majority of the tasks directly assessed the actions of the TC. A possible explanation is that TCs with high ability crewmen can distribute responsibility, have more confidence in their crew's competence, and therefore can better concentrate on their own job.

The Phase II Armor SPRP tests, reported here, demonstrated mental ability to be related to the C^2 performance of first-term and second-term soldiers. The Phase I Armor SPRP test showed mental ability to be related to simulated tank gunnery performance of soldiers during initial-entry training. Phase I analyses based on a Lanchester-type combat-attrition model indicated Cat IV soldiers performed at 73% of the level of Cat I&II soldiers[1]. Taken together, the Phase I & II tests have demonstrated mental ability effects over most of the domain of Armor tasks.

While the results show consistent effects of soldier quality, the utility of enlisting more costly quality soldiers could still be questioned. At the heart is the question of whether additional training can eliminate mental ability differences. The goal of Army training is to get all soldiers to perform to a set of standards. In theory, this approach is designed to eliminate the effects of individual differences. In practice, training is often structured to prepare a unit to perform well at an upcoming training exercise, e.g. Table VIII or NTC. Unit leaders are often able to identify and eliminate individual deficiencies for these training exercises through additional training or other safeguards. For example, the unit master gunner's job is to ensure that all tanks are boresighted before gunnery. If a crew cannot boresight, the master gunner will do it. In combat units will not have time to check up on weaker crews. If a crew cannot keep their tank boresighted, they will be killed or disabled.

The field test found a 30% difference between Cat I&II and Cat IV TCs' ability to prepare the tank for combat. If these tasks are not correctly performed, the technological advantages of the M1 tank are nullified. Plans are underway to place additional electronic equipment on future tanks, e.g., the Commander's Independent Thermal Viewer (CITV), and the InterVehicular Information System (IVIS). Based on the field test results, it is reasonable to expect that the new equipment will enhance higher mental ability crews performance more than lower mental ability crews. The result would be even bigger mental ability effects.

The modern battlefield will be dynamic and rife with uncertainty. AirLand Battle doctrine recognizes this and stresses the exploitation of the fluid battlefield. To do so, the Army needs soldiers who are resourceful and who can respond quickly to changing situations. Given a basic definition of mental ability as the ability to adapt to novel situations, AirLand Battle success is predicated on having quality soldiers who can respond to the dynamic battlefield, take on additional responsibilities, and make good decisions. While these points largely apply to leaders, enlisted armor crewmen will quickly be required to take on leadership responsibilities. After the first day of battle, units will be reconstituted. Some TCs will then become platoon leaders; gunners, loaders, and drivers will become TCs. The difference between being able to continue the battle and chaotic defeat will depend on how well the soldiers can handle the stress and take on the new responsibilities.

The SPRP test simulated stressful combat conditions using a third day of the war scenario. Crews were flown in, given an unfamiliar unit SOP which they had to quickly learn, made part of a reconstituted crew, and given a tank in need of maintenance. Throughout the field and SIMNET tests the crews encountered uncertainty, including the hideous simulated death of a crewman and platoon members. The results - roughly a 25% difference between Cat I&II and Cat IV crews. Given that the United States is investing $2.5 million with each tank it gives an Armor crew, a 25% decrement in performance is costly. By comparison, the cost to recruit and retain quality soldiers to obtain maximum weapon system effectiveness may be small.

The cumulative effects of mental category are even more dramatic when the SPRP findings are considered as combat multipliers. Consider the cumulative effects of the performance of Cat IV crewmen. Relative to the performance of Cat I&II crewmen, Cat IVs boresighted at 45%, hit targets at 73%[1], performed with a speed of 81%, effectively called for fire at 67%, and reported accurate grid coordinates at 55%. Furthermore, combat leaders will have greater confidence in quality crews, which will facilitate the execution of bold decisive actions. Given that combat is a series of battles in which these tasks must be performed over and over, the cumulative effects of mental ability will have a substantial impact on combat effectiveness. Higher quality soldiers equate to higher enemy attrition and higher unit survival.

References

1. Graham, S. E. (1989). Assessing the Impact of Mental Category on Simulated Tank Gunnery Performance. (ARI Research Report 1515). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

2. Tziner, A. & Eden, D. (1985). Effects of Tank Crew Composition on Tank Crew Performance: Does the Whole Equal the Sum of the Parts? Journal of Applied Psychology, 70.

3. Black, B. A. & Mitchell, K. J. (1986). Predicting Performance of M1 Gunners. (ARI Technical Report 707). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

4. Graham, S. E. (1987). Training Command, Control, and Communication Skills on SIMCAT. (ARI Research Report 1449). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

5. U.S. Army Armor Center (1989). SIMNET Users' Guide, Fort Knox, KY.

Identification of a Novel Cytokine-induced Effector Molecule
Involved in the Intracellular Destruction of
Leishmania, Franciscella, and *Plasmodium* sporozites:
Nitric Oxide (U)

* Shawn J. Green, CPT, MS, Anne H. Fortier, Dr., Robert M. Crawford, MAJ, MS,
Stephen L. Hoffman, CDR, Sylvie Mellouk, Dr., Monte S. Meltzer, COL, MC,
and Carol A. Nacy, Dr.
Walter Reed Army Institute of Research
Washington, D.C. 20307-5100

Introduction. Mammals, including man, synthesize nitrite (NO_2^-) and nitrate (NO_3^-) as a consequence of acute inflammatory responses. A number of different cells throughout the body produce these nitrogen oxides: endothelial cells that line blood and lymph vessels; hepatocytes, the parenchymal cells of the liver; microglia and other cells of brain.[1-3] The biochemical pathway for the production of NO_2^- and NO_3^- by nitrogen oxidation is known.[4] The terminal guanido nitrogen of the amino acid L-arginine is oxidized, and released as an unstable highly reactive intermediate, nitric oxide (NO). Citrulline is formed, and NO, with a half-life of milliseconds, is further oxidized to NO_2^- and NO_3^-. Citrulline, NO_2^-, and NO_3^-, then, are the byproducts of arginine degradation.

Despite its transitory nature, the reactive intermediate NO has striking biological effects on cells of the vascular, nervous, and immune systems. The production of exogenous NO has physiological importance as an endothelium relaxation factor.[5] Its clinical usefulness is readily apparent through the relief of angina by glyceryl trinitrate, a source of NO. This reactive intermediate is also strongly implicated as the effector molecule in the immunological destruction of tumor cells.[6] Of particular interest to our Department, whose mission is to develop strategies for nonspecific protection of soldiers exposed to hazardous biologic agents, was the report that white blood cells, specifically macrophages, are the source of NO_2^-/NO_3^- found during acute infections of mice with *Mycobacterium bovis* strain BCG.[7]

The macrophage is a pivotal cell in host response to any infectious agent. Not only does it provide a first line of defense, with its avid phagocytic and degradative capabilities, but it also plays a profoundly important role in the initiation of immune responses: it digests pathogens and presents the appropriate proteins (antigens) on its surface that stimulate the development of "antigen-specific" lymphocytes. These lymphocytes then orchestrate a complex series of reactions that are specifically tailored to eliminate the offending pathogen. As immunity develops, the macrophage acquires new, and expanded capabilities. It

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

now functions, at the direction of the lymphocytes, as a mobile cytotoxic cell that seeks out and destroys residual infectious agents. The most remarkable aspect of the macrophage is that its killing activities are totally nonspecific. The macrophage responds to signals sent by antigen-specific lymphocytes, and develops extraordinary killing mechanisms that are effective against any number of bacterial, parasitic, fungal, or viral pathogens. The macrophage does not discriminate one pathogen from another, and, thus eliminates not only the inciting pathogen that generated a specific immune response, but also any bystander infectious agents that happen to be present. This nonspecific killing by macrophages is the focus of our strategies for protection of soldiers from infectious challenge with agents for which there are presently no vaccines or chemotherapies, or with pathogens of unknown origin.

The process of macrophage activation is a excellent example of the complex intercellular communication network used by the immune system. The chemical signals that pass between the principal actors in any immune response (macrophages, T lymphocytes, B lymphocytes) are called cytokines. Cytokines are glycoproteins with molecular weights in the range of 10-70 kD, and are roughly equivalent to hormones of the endocrine system. They have one property that distinguishes them from true hormones, however: hormones are produced in one organ to affect cells at distant sites; cytokines are produced by cells locally at the site of an immune reaction, and are released in close proximity to the cell that is to receive the signal.

Cytokines can be thought of as the "words" of the immune system. Like any language, the more complex a concept, the more words are required to describe it. In immune reactions, this is particularly true for initiation of extraordinary immune effector mechanisms. We identified several different cytokine "phrases" and "sentences" that lymphocytes use to instruct macrophages to resist infection with pathogens.[8,9] Induction of intracellular killing is also very complex.[10,11] Several different cytokines induce the expression of intracellular killing activity by macrophages: an important activation signal is interferon γ (IFNγ), a cytokine released by immune T lymphocytes.[10,12]

The effector molecule(s) that activated macrophages actually use to eliminate pathogens has always been a mystery. We now report that synthesis of NO, the reactive intermediate formed by the oxidation of L-arginine, is required for microbicidal activity.[13] In this paper, we discuss the experimental approaches that allowed us to identify NO as an essential component of the intracellular killing mechanism of activated macrophages, and demonstrate the fascinating, but complex interactions of host cell, cytokines, and pathogens that regulate the production of NO by macrophages.

A model for immunotherapy: Intracellular destruction of *Leishmania major* by activated macrophages. Leishmaniasis is a spectrum of disfiguring tropical diseases that are of particular concern to the Army: there are no vaccines to protect troops assigned to endemic areas, and available chemotherapies are toxic and require patient hospitalization during administration. The etiologic agents of leishmanial disease are protozoan parasites which

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

infect and live only in macrophages. The diseases range from simple cutaneous lesions (caused by *L. major*, or *L. mexicana*) to fatal systemic infection (kala azar, caused by *L. donovani*). The leishmania parasites are under strong evolutionary pressure to subvert the normal macrophage antimicrobial effector activities for the sake of their own survival and propagation. The parasite successfully invade and multiply inside quiescent tissue macrophages.[10] Elimination of the parasite and resolution of disease must occur, then, by a change in the intracellular environment of macrophages from one that is supportive of parasite replication, to one that is hostile to survival. These diseases provide a unique opportunity to explore the potential of upregulating nonspecific macrophage intracellular killing activities for protection of soldiers at risk.

M. bovis strain BCG is a potent immunomodulating agent used throughout the world as a vaccine against tuberculosis. The immune reaction to BCG is so strong, and the organisms so difficult to eradicate, that nonspecific effector mechanisms of activated macrophages can be demonstrated for weeks or months after inoculation of the bacterium. We recognized the potential for testing BCG-induced nonspecific immunity with an intra-macrophage parasite, and developed several mouse models for disease caused by *L. major*.[14] Using these models, we demonstrated that BCG not only protected mice from cutaneous disease and lethal systemic infections, but these animals became solidly immune to rechallenge with the parasite.[15] Our success was based on learning how to focus the immune react-ion to BCG to the appropriate tissue: the skin, site of inoculation of the parasite.

Macrophages explanted from mice treated with BCG kill intracellular *L. major* amastigotes.[16] That BCG-activated macrophages produce NO_2^-/NO_3^- sugges-ted that NO may be the effector molecule for immune destruction of the intracellu-lar parasite, as it is for tumor cells.[6] We tested this possibility by treating mice with BCG, and explanting the activated peritoneal macrophages in vitro (Table 1).[13]

TABLE 1

Effects of N^GMMLA on macrophages from BCG-inoculated mice

N^GMMLA (mM):	Infected Macro-phages (% ± SEM):	Microbicidal Activity (%):	NO_2^- ($\mu M/72h$)
Macrophages from BCG-infected mice:			
0	6 ± 4	89	80
0.01	8 ± 3	85	67
0.10	46 ± 6	16	42
0.50	53 ± 4	4	16
Macrophages from untreated mice:			
0	55 ± 3	0	2

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

We infected BCG-activated macrophages with amastigotes of *L. major* in the presence or absence of N^G-monomethyl-L-arginine (N^GMMLA). N^GMMLA is widely used to study the nitrogen oxidation of L-arginine in a number of biological systems. It selectively inhibits nitrogen oxidation, but not other pathways of L-arginine metabolism (arginase activity or synthesis of polyamines) in cytotoxic macrophages.[17]

Both microbicidal activity and NO_2^- production, the stable endproduct of L-arginine-derived NO, were assessed at each concentration of the inhibitor. Macrophages activated in vivo with BCG killed 90% of intracellular parasites (Table 1). N^GMMLA, which interferes with the production of NO, blocked both NO_2^- production and intracellular killing activity by these activated cells: loss of one activity paralleled loss of the other.

The extraordinary abilities of BCG-activated macrophages to kill intracellular pathogens can be entirely reproduced by exposing normal macrophages to cytokines obtained from stimulated lymphocytes in vitro.[10] We recently identified IFNγ as an important cytokine for activation of these effector activities in both murine and human systems.[10,12] We also demonstrated in our mouse models the importance of IFNγ in the resolution of leishmanial disease. Genetically resistant mice treated with anti-IFNγ monoclonal antibody at the time of infection with *L. major* develop cutaneous lesions and systemic disease similar to that of genetically susceptible mice.[18] Since killing of *L. major* by BCG-activated macrophages correlates with the generation of NO, we examined whether IFNγ-treated macrophages also use NO as the effector molecule for destruction of intracellular parasites (Figure 1). Again, we used the specific inhibitor of L-arginine metabolism, N^GMMLA. Similar to activated macrophages obtained from BCG-infected mice, macrophages treated with IFNγ in vitro also developed impressive intracellular killing capability: >80% of macrophages infected with amastigotes cleared the intracellular parasites. Like that of BCG-activated macrophages, N^GMMLA inhibited this intracellular killing at concentrations >0.1 mM (Figure 1).

We confirmed the role of L-arginine-derived NO as an effector molecule for destruction of intracellular parasites by two different methods: we determined the specificity of N^GMMLA inhibition by adding excess L-arginine to the reaction mixture; and we depleted L-arginine in the culture medium by adding arginase. By either competition of the inhibitor with excess L-arginine (Table 2A) or enzyme digestion of L-arginine (Table 2B), we could demonstrate that L-arginine was essential for expression of microbicidal activity. Further, macrophage antileishmanial activity directly correlated with the metabolism of L-arginine to NO_2^-, the oxidative endproduct of the short-lived intermediate, NO (Table 3). The increase in NO_2^- production served as both a measurement of nitrogen oxidation of L-arginine by activated macrophages, and a quantitative indicator of macrophage activation and effector function.
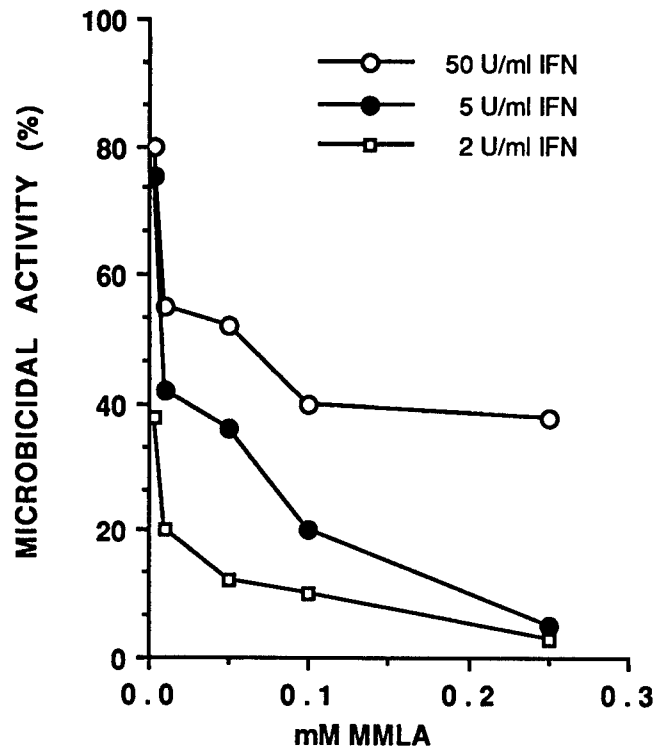
Figure 1. Effects of N^GMMLA on IFNγ stimulation of macrophage antileishmanial activites. 10^6 murine peritoneal macrophages were infected with amastigotes for 2 h, and then treated with different concentrations of IFNγ and N^GMMLA. After 3 days, cells were examined by light microscopy. Microbicidal activity was calculated by determining the change in percentage of infected cells in IFNγ-treated cultures compared to medium-treated cultures.

Although the biochemical pathway for nitrogen oxidation of L-arginine and its regulation are not fully characterized, it is known that NO inhibits selected iron-dependent enzymatic pathways in tumor targets.[19-22] NO from cytotoxic macrophages inhibits two oxidoreductases of the mitrochrondrial electron transport chain and the citric acid cycle enzyme aconitase in tumor cells. Each of the enzymes involved in these vital pathways has a catalytically active iron linked to a sulfur group that can be degraded by NO and released as a iron-nitrosyl complex.[19] We speculate that NO generated in IFNγ-activated macrophages may cause Fe-dependent enzyme inhibition of mitrochrondrial respiration in the *L. major* amastigote.

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

TABLE 2
L-arginine and expression of intracellular killing: addition of excess L-arginine[a] and arginase depletion of substrate for NO production[b]

Cells incubated 72 hr in:	Infected macro-phages (% ± SEM):	Microbicidal Activity (%):
A. *Addition of excess L-arginine:*		
medium	53 ± 6	
+ 0.1 mM MMLA	58 ± 4	0
20 U/ml IFNγ	3 ± 1	94
20 U/ml IFNγ		
+ 0.1 mM MMLA	35 ± 8	34
+ 0.1 mM MMLA + 1.4 mM L-arginine	8 ± 3	84
+ 2.4 mM L-arginine	2 ± 2	96
B. *Addition of arginase:*		
medium	65 ± 8	
+ 5 U/ml arginase	62 ± 6	5
20 U/ml IFNγ	10 ± 5	85
+ 5 U/ml arginase	43 ± 3	34

[a] Exogenous L-arginine was added to culture media at the time that *L. major*-infected macrophages were treated with IFNγ and N^GMMLA.

[b] Arginase was added to culture medium prior to use in the assay.

TABLE 3
Correlation between NO_2^- production and microbicidal activity by IFNγ-stimulated macrophages[a]

Cells treated 72 hr with:	Microbicidal Activity (%):	NO_2^- (uM/10^6cells/72 h)
0 U/ml IFNγ	0	1 ± 2
10 U/ml IFNγ	90	65 ± 11
+ 1 mM N^GMMLA	4	1 ± 0
+ 5 U/ml Arginase	30	8 ± 5

[a] After 72 hrs, culture fluids were assayed for NO_2^- by the Greiss reaction to form a chromophore; absorbance was measured at 543 nm. Cell smears were examined by light microscopy for determination of microbicidal activity.

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

Parasite participation in the stimulation of NO production by activated macrophages. Unlike other macrophage effector activities, the initiation of nitrogen oxidation is a multisignal event.[13,24] The level of nitrogen oxidation after stimulation of uninfected macrophages with IFNγ alone is not sufficient to affect the viability of intracellular amastigotes.[13] In over 40 experiments, macrophages exposed to IFNγ alone produced an average of 2 ± 2 mM NO_2^-/72 h. By correlating intracellular killing activities and NO_2^- accumulation in these 40 experiments, we determined that 30 ± 5 uM NO_2^-/72 h reflected the level of intracellular NO required to kill the parasite. The paradox we encountered in the *Leishmania* studies is that the cells which have potent microbicidal activity and high levels of nitrite were treated with a single activating agent, IFNγ. How then does the IFNγ activated macrophage receive a second signal for intracellular killing of *L. major* amastigotes?

Microbial products, such as the bacterial immunostimulants lipopolysacchride (LPS) or muramyl dipeptide (MDP), are known to induce nitrogen oxidation in combination with IFNγ.[23,24] In a similar fashion, we found that amastigotes themselves provide an effective second signal for NO production (Table 4).[24]

TABLE 4
Second signals and the induction of NO_2^- synthesis[a]

Cells treated with:			NO_2^- Production by: (uM/72hrs)	
IFNγ (10U/ml)	LPS (10ng/ml)	Amastigotes (10^6/ml)	*Peritoneal Macrophages*	*Bone marrow Macrophages*
-	-	-	0	0
+	-	-	2 ± 2	0
-	+	-	1 ± 1	8 ± 1
+	+	-	71 ± 8	60 ± 4
-	-	+	1 ± 1	0
+	-	+	64 ± 5^b	58 ± 1^b

[a] 10^6 cells were exposed to different activating agents; culture fluids were assayed for nitrites after 72 h by the Greiss reaction. Results are expressed as mean NO_2^- production $\pm$ SEM for 2 samples of each treatment group.

[b] Amastigote infected peritoneal and bone marrow-derived macrophages treated with IFNγ had >90% microbicidal activity.

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

The irony of this observation is clear: by providing a signal for the initiation of the L-arginine dependent killing mechanisms in IFNγ stimulated macrophages, the amastigotes participate in their own destruction.

There is always some concern in assessing macrophage function in vitro, because contaminating bacterial LPS may confound the results of any activation scheme. Since LPS is such a potent second signal for NO production by IFNγ treated cells, we were particularly interested in ruling out contaminating LPS as the actual signal delivered by the amastigote preparation.[24] To characterize amastigote initiation of NO synthesis, we: (a) checked the level of LPS in amastigote preparations by Limulus amoebocyte lysate assay. Levels of LPS, if present, were below 3ng/ml, the lower limits of the assay. LPS dose responses suggested that >50 ng LPS is required to initiate NO synthesis; (b) performed the assays in the presence of polymixin B, an antibiotic that binds to the lipid A region of LPS. Polymixin B inhibited completely the triggering of NO by LPS, but had absolutely no effect on triggering of NO by amastigotes; and (c) treated the two trigger signals, LPS and parasites, at 80°C for 5 min. Amastigote activity was totally abolished, while LPS, which is remarkably heat resistant, was unaffected. We conclude that the parasites and LPS provide two distinct second signals.

Muliple signal control of macrophage activation has been studied for a number of years. The entire process is very tightly regulated for the expression of these extraordinary effector reactions[25,27]: the macrophage must be in a receptive state to receive activation signals, the signals must be presented in the appropriate time frame, and the signals must be presented in the correct sequence. At each of these steps, the immune system has designed bail-out options to control the process. Therefore, if cells are in the wrong differentiative state, or if the signals are not received before the transient receptive state is concluded, or if the cells receive the signals out of sequence, the whole process is aborted. Why might the body so carefully regulate these effector reactions? Most likely, this tight regulation has evolved to limit the damage to normal tissues by NO. In fact, NO that is generated for the destruction of tumor targets and microorganisms also has profound effects on even the effector cell that generated the NO, the macrophage.[27] Therefore, the limitation of NO production to cells that have actually encountered a pathogen makes teleological sense: only cells that have a pathogen to kill will generate the NO, and the amount of NO leakage into surrounding normal tissues will be minimized. For the *Leishmania*, the target microorganism itself provides the second signal. We also found that other intracellular parasites that grow in unstimulated macrophages, such as metacyclic *L. major* promastigotes, *Toxoplasma gondi*, and *Francisella tularensis* were all susceptible to the lethal effects of NO, and were all able to trigger NO_2^- synthesis in macrophages treated solely with IFNγ.

The nature of the second signal provided by intracellular targets is a fascinating question. In the case of gram negative bacteria, LPS is the most likely candidate. For other non-bacterial parasites, the signal is unknown, but appears from our studies with *L. major* to be independent of LPS.

In an effort to identify the cellular consequences of infection with *L. major*, we found that amastigotes trigger the synthesis and release of tumor

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

necrosis factor α (TNFα), a cytokine with pleiotrophic effects on a wide variety cell including macrophages.[24,31] Of all the cytokines previously tested for macrophage activation in vitro, only exogenously added TNFα, in combination with IFNγ, triggered nitrogen oxidation of L-arginine.[23,24,28] To test whether TNFα was involved in the amastigote initiation of NO_2^- production, we infected macrophages in the presence of anti-TNFα antibodies or control antibodies of irrelevant specificity, and then treated cells with IFNγ. Anti-TNFα antibody, but not the control antibody, totally abrogated NO_2^- production as well as microbicidal activity (Table 5). The addition of excess recombinant TNFα to the antibody-treated, amastigote-infected macrophages exposed to IFNγ restored their ability to produce nitrite and kill parasites. Amastigotes, then, triggered the synthesis and

TABLE 5

Effects of antiTNFα antibodies on intracellular killing, nitrite production, and TNFα secretion

Culture Conditions:	Microbicidal Activity (%)[b]	NO_2^{-}[b] (μM/ml;70 h)	TNFα[a] (pg/ml)
macrophages:		0	0
+ amastigotes		0	20
+ amastigotes +IFNγ:	97 $\pm$ 6	40 $\pm$ 9	60
+ antiTNFα	3 $\pm$ 3	7 $\pm$ 4	<2
+ antiTNFα + (TNFα @ 3h)[c]	87 $\pm$ 4	41 $\pm$ 9	- -
+ control MAb	91 $\pm$ 0	42 $\pm$ 5	76

[a] 10^6 resident peritoneal macrophages were treated with 50 μg/ml anti-TNFα Ab or control IgG MAb, followed by the addition of amastigotes and 10 U/ml IFNγ. After 72 h, microbicidal activity and NO_2^- production were measured.

[b] Macrophage culture medium was collected after a 2 h exposure to amastigotes. TNFα activity was titrated on actinomycin D-treated, TNFα-sensitive L-929 cells. After 20 h, L-292 cell viability was assessed by the crystal violet-colormetric assay and TNFα titers were determined.[25]

[c] 100 U/ml TNFα was added 3 h after the addition of the antiTNFα Ab.

release of TNFα, which in turn acted as an autocrine factor to induce NO-dependent killing activity by IFNγ-primed macrophages. We envision the interaction of cells, cytokines, and parasites to be that depicted in Figure 2.

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY
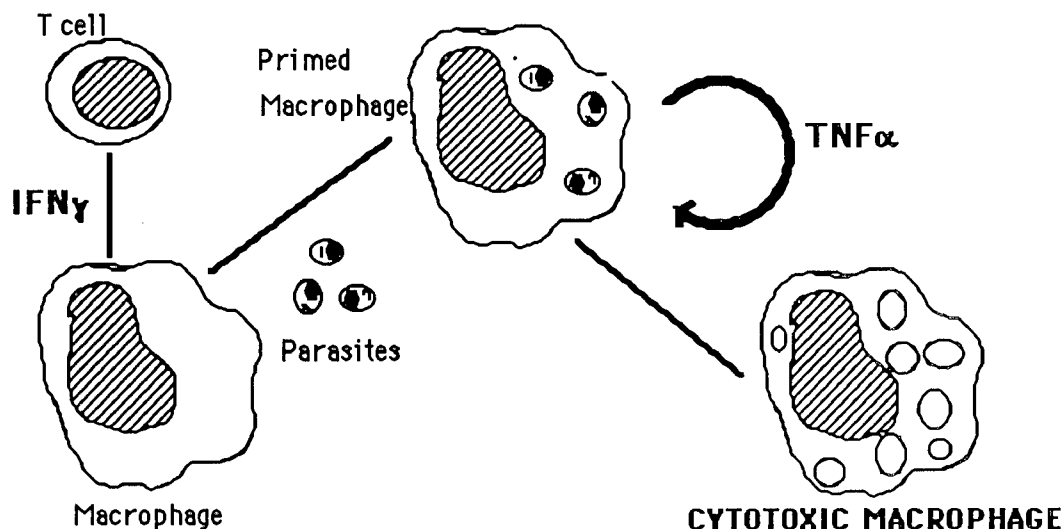
Figure 2. Autocrine regulation of macrophage intracellular killing activities.

Macrophage-derived NO: A defense against biological threat agents. An overriding concern of the Department of Army is the management of respiratory infections, particularly those initiated by pathogens that pose a biological warfare threat. The theoretical possibility of bioengineered organisms combining pathogenic traits from many different agents suggests that specific vaccination will be of limited use. The best possible host defense in instances where the pathogen is not identified (or identifiable) is nonspecific upregulation of the immune system.

We recently developed a program to characterize host response to the potential biological warfare agent, *Francisella tularensis*, an organism that has already been weaponized. The goal of this program is to develop strategies for nonspecific protection of soldiers exposed to aerosolized pathogens, and *F. tularensis* strain LVS is used as a model. Again, we found that the principle target and first line of defense against *Francisella* is the macrophage. To study the interaction of this pathogen with host cells, we developed in vitro assays for detection of bacterial growth as an assessment of viable organisms inside infected macrophages. As shown in Table 6, *F. tularensis* grows well in unstimulated macrophages: the input inoculum was $10^{3.5}$, and 48 h later we could recover 10^8 bacteria. IFNγ-treated macrophages exposed to *F. tularensis* effectively blocked replication of this pathogen. Furthermore, IFNγ-induced cytostasis was suppressed by blocking nitrogen oxidation of L-arginine with N^GMMLA. As described for *L. major*, the induction of NO synthesis by IFNγ-activated macro-

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

TABLE 6
Effects of N^G MMLA on IFNγ Activation of Macrophages
for Suppression of the Growth of *F. tularensis* [a]

Culture Conditions:	Bacterial Growth (CFU at 48 hrs):	NO_2^- production ($\mu M/48$ h)
Macrophages infected with F. tularensis and treated with:		
+ Medium	1×10^8	2 ± 1
+ 10 U/ml IFNγ	5×10^3	18 ± 2
+ 10 U/ml IFNγ + 0.1 mM N^GMMLA	2×10^8	1 ± 1
+ 0.1 mM MMLA	1×10^8	0

[a] Macrophages infected with *F. tularensis* and treated with IFNγ, with or without the inhibitor, were disrupted by sonication. The bacteria were plated in serial 10-fold dilutions into 96 well plates with bacterial growth medium. Turbidity, a measure of bacterial replication, was assessed macroscopically, and titers were determined by the reciprocol of the last dilution to show turbidity.

phage was triggered by the target itself. In this case, the likely signal is LPS derived from the cell wall of *F. tularensis*.

Exoerythrocytic stages of *Plasmodium bergheii*: Hepatocyte NO implicated as effector molecule for IFNγ-induced killing events. There are several reports that document a role for IFNγ in control of malaria sporozoite invasion and replication in liver cells.[29,30] The target host cell for the parasite in liver is the hepatocyte. The mechanisms involved in the destruction of these parasites has never been resolved. Since hepatocytes produce NO as a consequence of IFNγ stimulation,[31] we analysed whether N^GMMLA would affect hepatocyte destruction of sporozoites in vitro (Table 7). N^GMMLA partially reversed the IFNγ-induced inhibition of sporozoite invasion. These preliminary results suggest that NO may also be involved in the IFNγ-induced killing of this protozoan parasite by a cell not normally involved in immune reactions.

NO, the twentieth century magic bullet. Intracellular pathogens, sequestered from the extracellular environment, are notoriously difficult to eradicate by humoral host defenses (ie., specfic antibody) and drug therapies.

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

TABLE 7
Effects of N^GMMLA on IFNγ Activation of Hepatocytes
for Inhibition of Invasion of *P. bergheii* sporozoites[a]

Culture Conditions:	Sporozoites/10^6 Hepatocytes:	Decrease in invasion (%)
Hepatocytes exposed to sporozoites and treated with:		
+ Medium	1766	0
+ 1000 U/ml IFNγ	715	60
+ 1000 U/ml IFNγ + 0.1 mM N^GMMLA	1282	27
+ 0.1 mM MMLA	1787	0

[a] Invasion was estimated by microscopic examination of successive fields by phase microscopy.

Drug therapies are confounded by the issues of achieving effective drug levels not only in tissues, but inside individual infected cells. It is nothing short of remarkable that one single effector molecule, NO, is capable of killing such a diversity of pathogenic agents. The agents now shown to be killed by NO include bacteria (*Francisella tularensis*, this study), parasites (*Leishmania major*, this study; *Schistosoma mansoni*[32]), fungi[17], and neoplastic cells.[19-24] Further, NO is produced not only by cells of the immune system, but also a variety of other cells types in the body.[1-3] Thus, it is possible to envision the upregulation of NO synthesis as an antidote to pathogenic agents sequestered in all sorts of host cells throughout the body.

The tight regulation of NO synthesis, with multiple endogenous and autocrine cytokines required to initiate intracellular killing functions, suggests that indiscriminate upregulation of NO may be injurious to the host. It is a reported that NO synthesized by activated macrophages affects mitochondrial respiration not only of the tumor target, but also of the macrophage.[27] Although repair processes in the macrophage appear to be more efficient than in tumor targets, downregulation of this pathway is equally as important as its initation. Therefore, the macrophage has developed strategies to protect the host from toxic NO and related N-nitrosating compounds which are carcinogenic. For example, the release of arginase by cytotoxic macrophages is believed to downregulate NO production by depleting extracellular arginine, thus deny the activated macrophage of the sole substrate for this effector pathway.[17] Recently, we discovered that the pretreatment of macrophages with the either transforming

GREEN, FORTIER, CRAWFORD, HOFFMAN, MELLOUK, MELTZER & NACY

growth factor β or interleukin 4, potent immunoregulatory cytokines, suppresses NO production as well as antimicrobicidal activities induced by the combination of IFNγ and TNFα. It is apparent that the complexity of regulating nonspecific immunity is reflected in the "complex sentences" created by different "cytokine words" that are used to instruct macrophages to protect us against both infection and the effector activity itself.

The recognition of NO as a common effector molecule for destruction of physiologically and phylogenetically diverse pathogens of interest to the Army is but the first step in a series of studies whose aim is the nonspecific protection of soldiers from infectious agents. The design or identification of drugs that can selectively regulate this effector molecule could prove useful in conventional infectious diseases, as well as biological warfare threats.

REFERENCES

1. Palmer, R.J.M., Ashton, D.S., and Moncada, S. (1988) *Nature* 333:664.
2. Vallance, P., Collier, J., and Moncada, S. (1989) *Lancet*, 328:997.
3. Garthwaite, J. Charles, S.L., and Chess-Williams, R. (1988) *Nature* 336:385.
4. Marletta, M.A. (1989) *Chem. Res. Toxicol.* 1:249.
5. Palmer, R.J.M., Ferrige, A.G., and Moncada, S. (1989) *Nature* 327:524.
6. Hibbs, J.B., Jr., Taintor, R.B., and Vavrin, Z. (1987) *Science* 235:473.
7. Stuehr, D.J., and Marletta, M.A. (1987) *J. Immunol.* 139:519.
8. Davis, C.E., Belosevic, M., Meltzer, M.S., and Nacy, C.A. (1988) *J. Immunol.* 141:627.
9. Belosevic, M., Davis, C.E., Meltzer, M.S., and Nacy, C.A. (1988) *J. Immunol.* 141:890.
10. Nacy, C.A., Meltzer, M.S., Leonard, E.J., and Wyler, D.J. (1981) *J. Immunol.* 127:2381.
11. Nacy, C.A., A.H. Fortier, Meltzer, M.S, Buchmeier, N.A., Schreiber, R.D. (1985) *J. Immunol.* 135:3505.
12. Hoover, D.L., Finbloom, D.S., Crawford, R.M., Nacy, C.A., Gilbreath, M., and Meltzer, M.S. (1986) *J. Immunol.* 136:1395.
13. Green, S.J., Meltzer, M.S., Hibbs, J.B., Jr., and Nacy, C.A. (1990) *J. Immunol.* 144:278.
14. Nacy, C.A., Fortier, A.H., Pappas, M.G., and Henry, R.R. (1983) *Cell. Immunol.* 77:298.
15. Fortier, A.H., Mock, B.A., Meltzer, M.S., and Nacy, C.A. (1987) *Infect. Immunity* 55:1714.
16. Papas, M.G. and Nacy, C.A. (1983) *Cell. Immunol.* 80:217.
17. Granger, D.L., Hibbs, J.B., Perfect, J.R., and Durack, D.T. (1990) *J. Clin. Invest.* 85:264.
18. Belosevic, M., Finbloom, D.S., van der Meide, P., Slayter, M.V., and Nacy, C.A. (1989) *J. Immunol.* 143:266.
19. Hibbs, J.B., Taintor, R. Vavrin, Z., and Raclin, E. (1988) *Biophys. Res. Comm.* 157:87.
21. Drapier, J. and Hibbs, J.B. (1988) *J. Immunol.* 140:2829.

22. Hibbs, J.B., Vavrin, Z., and Taintor, R.R. (1987) *J. Immunol.* 138:550.
23. Drapier, J., Wietzerbin, J., and Hibbs, J.B. (1988) *Eur. J. Immunol.* 18:1587.
24. Green, S.J., Crawford, R.M., Hocksmeyer, J., Hibbs, J.B., Jr., Meltzer, M.S., and Nacy, C.A. (1990) *J. Immunol.,* in press.
25. Ruco, L.P. and Meltzer, M.S. (1978) *Cell. Immunol.* 41:35.
26. Ruco, L.P. and Meltzer, M.S. (1978) *J. Immunol.* 121:2035.
27. Amber, I.J., Hibbs, J.B., Jr., Taintor, R,R, and Vavrin, Z. (1988) *J. Leukocyte Bio.* 43:187.
28. Ding, A.H., Nathan, C.F., and Stuehr, D.J. (1988) *J. Immunol.*141:2407.
29. Mellouk, S., Maheshwari, R.K, Rhodes-Feuillette, A., Beaudoin, R.L., Berbiguier, N., Matile, H., Miltgen, F., Landau, I., Pied, S., Chigot, J.P., Friedman, R.M., and Mazier, D. (1987) *J. Immunol.* 139:4192.
30. Ferreira, A., Schofield, L., Ena, V., Schellekens, H., VanDer Meide, P., Collins, W.E., Nussenzweig, R.S., and Nussenzweig, V. (1986) *Science* 232:881.
31. Curran, R.D., Billiar, T.R., Stuehr, D.J., Hofmann, K., and Simmons, R.L. (1989) *J. Exp. Med.* 170:1769.
32. James, S.L. and Glavin, J. (1989) *J. Immunol.* 143:4208.

GRILLO & HUDSON

Protection from Ricin Toxicity by Brefeldin A (U)

*F. Gregory Grillo, MAJ
Thomas Hudson, Dr.
Walter Reed Army Institute of Research
Washington, DC. 20307-5100

Ricin is a heterodimeric, proteinaceous, toxin extracted from castor oil beans which kills cells by disrupting protein synthesis. The lethal mechanism appears to be the destruction of critical nucleotide residues in 28S rRNA which inhibits ribosome function.[1] This modification results from the enzymatic activity of the toxin's alpha subunit, an N-glycosidase. The means by which alpha is delivered to its intracellular target is less clear. Binding of ricin to the cell is the province of the beta subunit which possesses two sites with high affinity for galactose residues on surface glycoproteins and glycolipids. The role of beta may not end there, but following binding the toxin is endocytosed. How ricin is translocated from the endocytotic vesicle for its fateful meeting with ribososmes remains a mystery.

The possibility that Golgi structures may be involved in toxin routing is suggested inferentially, by evidence that monensin, an agent known to perturb Golgi function, can modify intoxication[2], and directly, by electron micrographic studies which show labeled ricin associated with the trans-Golgi network.[3] We decided to test whether retrograde passage through the Golgi (i.e. transfer from the trans Golgi network through the trans, medial, and cis Golgi stacks) for final delivery to ribosomes in the endoplasmic reticulum is necessary for ricin toxicity by employing a newly described inhibitor, Brefeldin A (BFA).

This compound is a macrolide antibiotic product of the fungus Eupenicillium brefeldianum which results in the dissolution of cis Golgi stacks with a redistribution of the enzymic residents of these structures to the endoplasmic reticulum.[4,5,6] Brefeldin A is well tolerated by numerous cell lines. It inhibits secretion of glycoproteins which undergo Golgi processing but has no independent effects on cellular energetics or protein synthesis. The Jurkat cell line (derived from a human T cell leukemia) was selected to study the

GRILLO & HUDSON

effects of BFA on ricin toxicity because of its easy culturability and sensitivity to the toxin.

To assure that BFA has effects on Jurkat cells similar to those reported in the literature for other cell types, we first demonstrated that the compound inhibited both constitutive and conconavalin A stimulated, IL-2 secretion by these cells (data not shown). In addition, electron microscopy demonstrated the dissolution of Golgi structures by BFA.

In experiments designed to screen for protection, cells were preincubated with ricin at 4° C for 1 hr and then washed free from unbound toxin. They were then incubated at 37° C in media plus or minus BFA. At appropriate intervals aliquots were withdrawn to assay protein synthesis by 1 hr pulse exposures to ^{14}C leucine followed by collection of acid precipitated protein on glass fiber filters and scintillation counting. (This protocol was selected to minimize potential complicating effects of BFA on ricin binding.) As shown in Fig. 1, treatment with BFA resulted in 100% protection of cells within the time course shown and at least 8 additional hrs (data not presented).

To further quantitate the protection afforded by BFA, we next tested cells exposed to a wide range of doses of each agent. As shown in Fig. 2, the I_{50} for ricin was 10^{-11} M for control cells and 10^{-7} M for cells treated with BFA. Complete protection from 10^{-9} M ricin was achieved by as little as 0.5ug/ml BFA (Fig. 3).

Since ricin has emerged as a potential biological threat agent and protective strategies need to be designed, the kinetics of BFA protection from ricin toxicity were examined. As shown in Fig. 4, delay of BFA exposure to ricin treated cells by as little as 20 minutes resulted in a loss of 50% of the protection. After 30 minutes delay, the agent would be of no apparent value.

An additional practical concern involves how long BFA treatment needs to be continued to permanently protect cells doomed by ricin binding. This was addressed in the experiment in Fig. 5. Here, a 1 hr BFA pulse was given after ricin binding. The BFA was then washed away, but cells remained protected from intoxication at the 50-60% level 4 additional hrs.

BFA has no direct effect on the enzymatic activity of ricin alpha chain in an _in vitro_ assay where depurination of 28S RNA was measured (data not shown).

In summary, BFA affords remarkable protection from ricin toxicity in the Jurkat lymphocyte cell line. Mechanistically, ricin delivery, which appears to require retrograde Golgi passage, is the presumed target of BFA action. These data suggest a multitude of strategies to protect individuals at risk for contact with toxin as well as for those who have already been exposed to the agent.

GRILLO & HUDSON
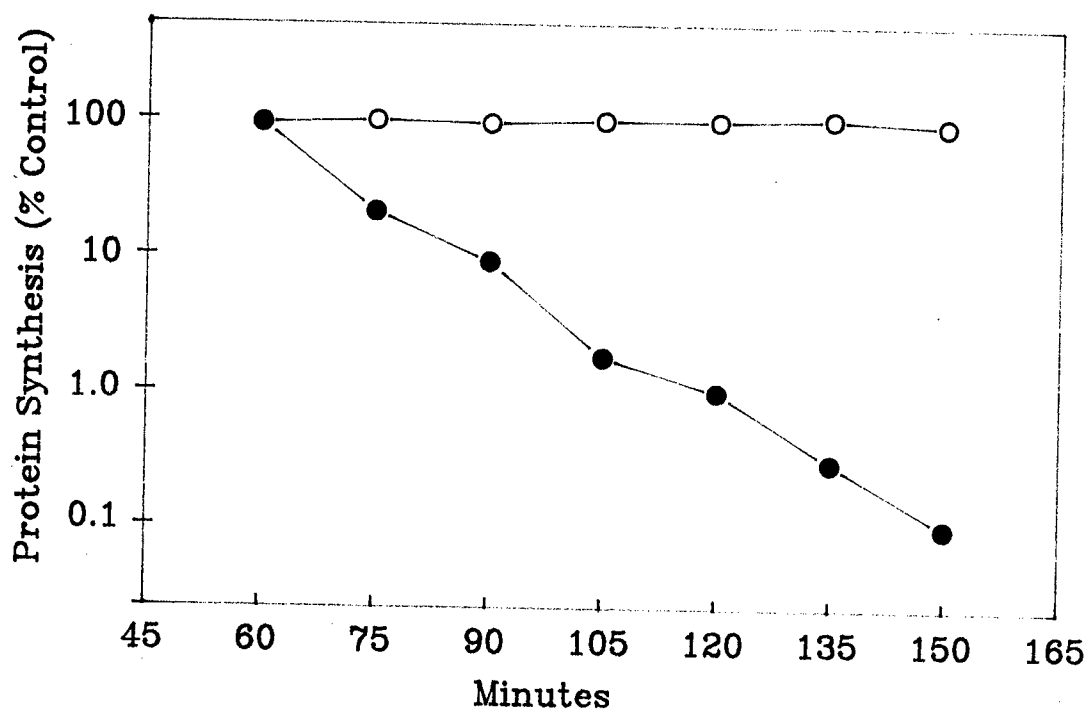
FIG. 1 TIME COURSE OF BFA PROTECTION

Cells were exposed to ricin (10^{-9}M) for 1 hr at 4° C, washed, and returned to control media (closed circles) or media containing 10 ug/ml BFA (open circles). They were then incubated at 37° C for the appropriate time and aliquots were withdrawn for 1 hr pulse exposures to 0.1 uCi ^{14}C leucine. Protein was acid precipitated and collected on glass fiber filters for scintillation counting. Control studies in the absence of ricin also were done. Data expressed as percent control protein synthesis were calculated compared to their own control. Protein synthesis inhibition by BFA alone ranged from 0-20%. Data are from a representative experiment performed in triplicate.
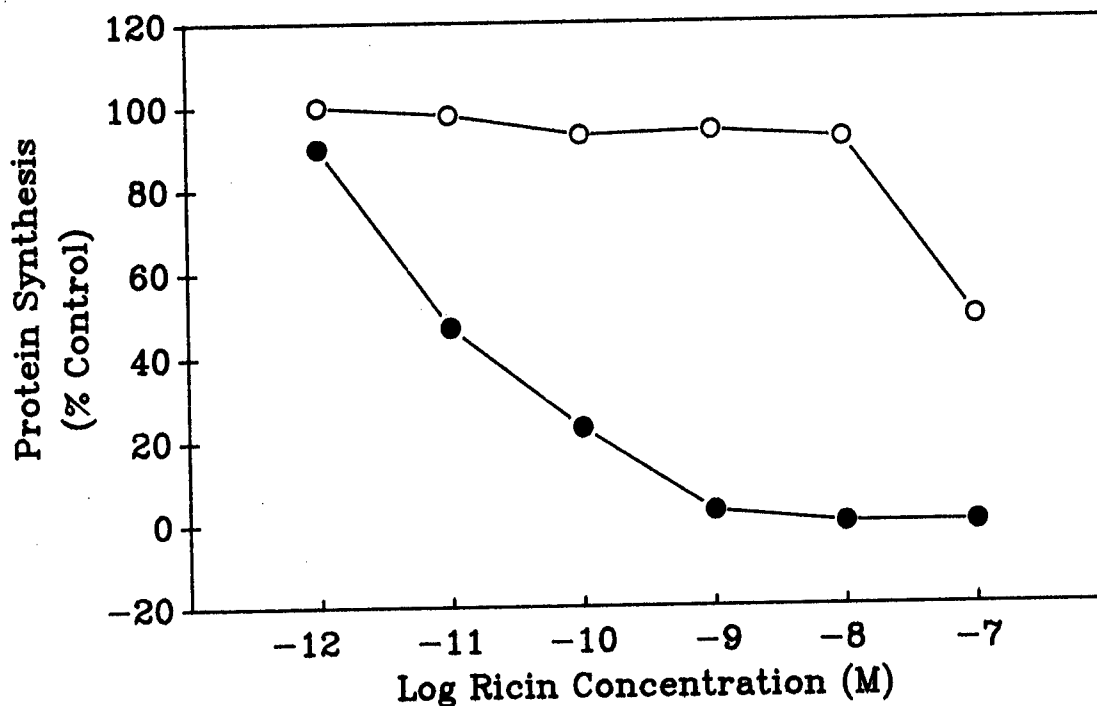
FIG. 2   RICIN DOSE-RESPONSE +/- BFA

Cells were exposed to ricin in appropriate doses as in Fig. 1.
A 3 hr incubation in control media (closed circles) or 10 ug/ml BFA
(open circles) was followed by 1 hr pulse treatments with ^{14}C leucine
as descibed above. Percent control protein synthesis was determined
as above. Data are from a representative experiment performed in
triplicate.

GRILLO & HUDSON
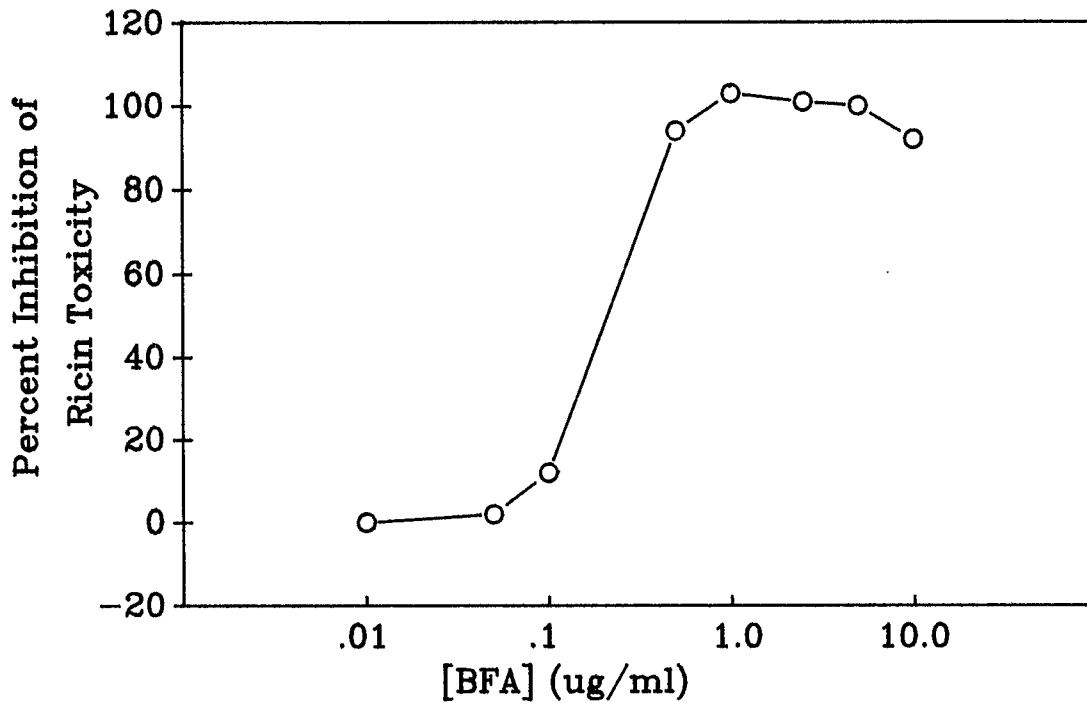
FIG. 3 RICIN TOXICITY DEPENDENCY ON BFA DOSE

Experiment was performed with ricin at 10^{-9} M as in Fig. 2. Percent inhibition of ricin toxicity was determined from percent control protein synthesis data calculated as above. Data shown are from a representative experiment performed in triplicate.

GRILLO & HUDSON
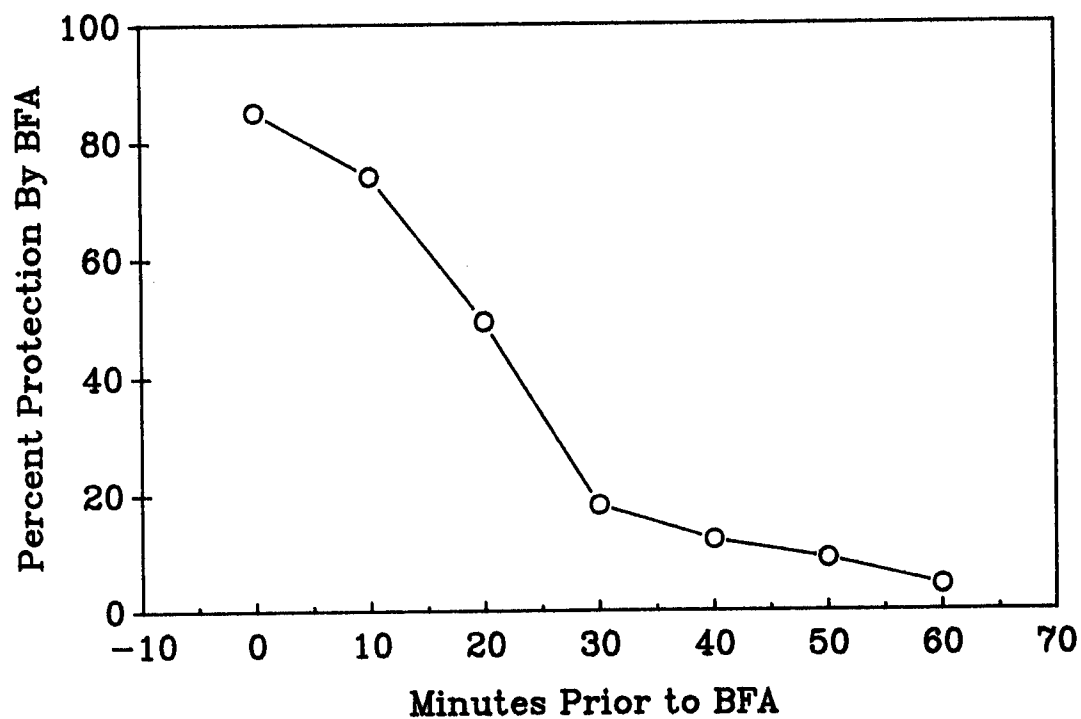
FIG. 4 BFA PROTECTION FOLLOWING RICIN EXPOSURE

Cells were exposed to 10^{-10} M ricin as described above and then were incubated in control media for the appropriate time prior to placement in 5 ug/ml BFA for 1 hr pulses with ^{14}C leucine. Data are derived from percent control protein synthesis calculated as above. Data shown are the mean of 3 experiments performed in triplicate.

GRILLO & HUDSON
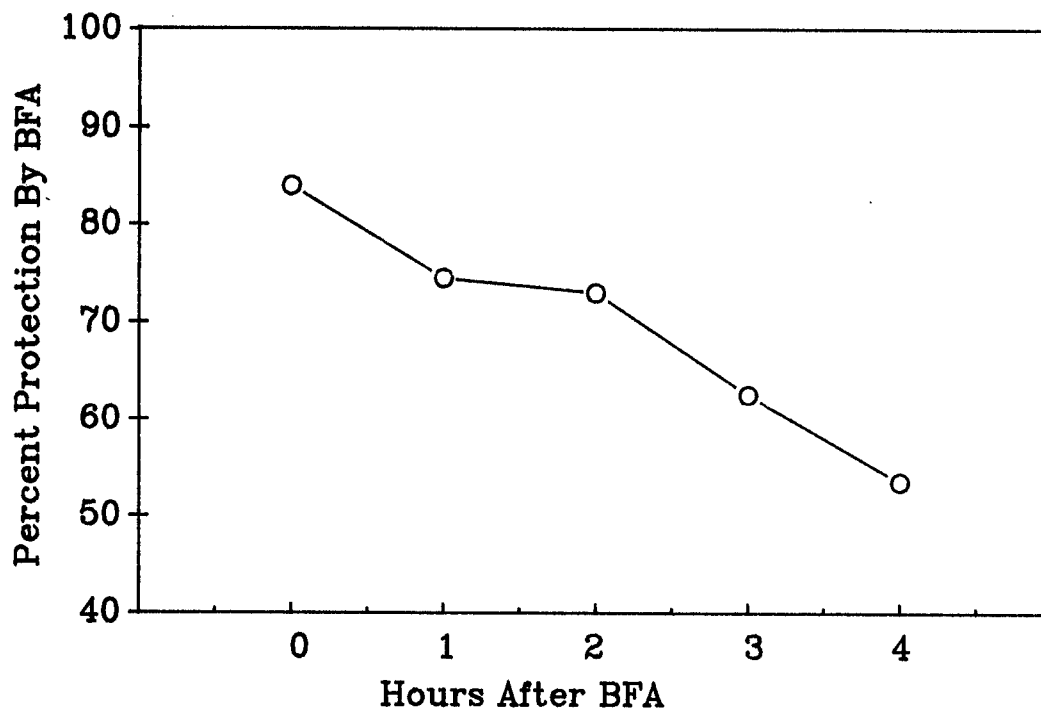
FIG. 5 PROTECTION BY BFA PULSE TREATMENTS

Cells exposed to 10^{-10} M ricin were subsequently exposed to a 1 hr pulse of BFA (5ug/ml). They were then washed and allowed to incubate for an appropriate time prior to a 1 hr pulse exposure to ^{14}C leucine. Data are derived as above and are the mean of 3 experiments performed in triplicate.

GRILLO & HUDSON

REFERENCES

1. Endo, Y., Mitsui, K., Motizuki, M., and Tsurugi, K., (1987) J. Biol. Chem. 262, 5908-5912.

2. Youle, R.J. and Colombatti, M., (1987) J. Biol. Chem. 262, 4676-4682.

3. van Deurs, B., Sandvig, K., Peterson, O.P., Olsnes, S., Simons, K., and Griffiths, G., (1988) J. Cell Biol. 106, 253-267.

4. Doms, R.W., Russ, G., and Yewdell, J.W., (1989) J. Cell Biol. 109, 61-72.

5. Lippincott-Schwartz, J., Yuan, L.C., Bonifacino, J.S., and Klausner, R.D., (1989) Cell 56, 801-813.

6. Fujiwara, T., Oda, K., Yokata, S., Takatsuki, A., and Ikehara, Y., (1988) J. Biol. Chem. 263, 18545-18552.

GROWER, ABUEME & SENG

Development of Topically Applied Anti-Inflammatory Agents for
Treating Acute Pulpal Inflammation in Combat Situations (U)

*Marvin F. Grower, COL, DC
Jeremias M. Abueme, SPC
George F. Seng, COL, DC
US Army Institute of Dental Research
Washington, DC 20307-5300

INTRODUCTION

A large percentage of patients seeking emergency treatment for
non-traumatic pain of dental origin are manifesting the symptoms of "acute
pulpitits" which is due mainly to caries (Trowbridge, 1985). Such acute
painful pulpitis is the inevitable sequelle of untreated dental caries and
concomitant chronic progressive pulpal inflammation (Ingle et al 1976;
Simon, 1980; Seltzer & Bender, 1984). In decayed human teeth, bacterially
produced acids demineralize dentinal tubules to produce affected dentin.
The acids and other metabolic by products are then able to transverse the
enlarged tubules into the pulp to produce tissue damage and inflammation
(pulpitis). The pain produced by this process is due to pressure on
sensitized pulpal pain receptors and is in turn exacerbated by pressure
induced circulatory stasis (Seltzer & Bender, 1984). In the military this
dental problem is a significant factor in limiting combat readiness. For
instance in Vietnam 1 out of every 7 soldiers was lost for 3 duty days
every year due to non-combat related dental emergencies (Cassidy, 1970).

The initial stages of acute pulpal inflammation, prior to actual
bacterial invasion or pulpal death, may be reversed and the pulp saved if
prompt and effective local measures are utilized to reduce the inflammatory
process. Such treatments require local application since the pressure
induced circulatory stasis (Seltzer & Bender, 1984) prevents systemically
administered agents from entering the pulp chamber to exert
anti-inflammatory effects. The procedures currently utilized to reduce
pulpal inflammation involve the removal of most of the decayed dentin by a
skilled professional and the application of a sedative-filling; however
agents presently used, such as zinc oxide and eugenol are of limited value
as they serve only as obtundants. They do not directly affect the
inflammatory process and do not predictably prevent pulpal death and the
need for more extensive endodontic or surgical intervention.

The non-steroidal anti-inflammatory drugs such as Indomethacin appear
to possess several properties which suggest that they might represent a new
and effective treatment of caries related pulpal inflammation. They act

GROWER, ABUEME & SENG

directly to reduce inflammation and circulatory stasis (Ferriera & Vane, 1974), they relieve pain through a local topical effect (Huskisson, 1983), they do not interfere with tissue healing (Alvarez et al, 1984), and they are bacteriostatic in nature (Schwartz & Mandel, 1972).

In order to exert a local effect on pulpal inflammation, topically applied Indomethacin must be transported through a layer of dentin. This would probably utilize a multistep process similar to that observed for drug penetration through skin (Ogiso et al, 1989). The absorption of Indomethacin through dentin would thus include transfer of the drug from the vehicle to the dentin, diffusive penetration of the drug through the dentinal tubules, and subsequent removal via the pulpal fluid and vasulature.

In the initial stages of this project (Seng et al, 1989) a model system was developed which approximated both the physical properties and physiologic characteristics of a carious tooth with an inflammed pulp. The system volumetrically approximated the natural tooth and was used to produce dentin similar in structure/composition to the demineralized carious dentin observed in vivo. It also allowed for the gradual acidic shifts in pulpal pH that would be encountered in vivo in inflamed tissue (Brune et al, 1976).

The work reported in this paper utilized this in vitro tooth system to determine potential vehicles for delivering therapeutic concentrations of Indomethacin through demineralized "caries like" and undecalcified dentin into the pulpal chamber within a 24 to 48 hour time period. The agents thus developed could then be used for subsequent in vivo testing of Indomethacin in preventing pulp death by relieving the inflammation and edema of an acute pulpitis.

MATERIALS AND METHODS

Delrin plastic chambers designed to structurally and qualitatively resemble human teeth (Fig. 1) as described by Seng et al (1989), were used both to decalcify dentin discs as a model of carious dentin, and to measure the penetration of Indomethacin through decalcified and undecalcified dentin. Freshly extracted non-carious human molar teeth which were stored in pH 7.6, 0.01M phosphate buffered saline (PBS) containing 0.1% sodium azide as an antimicrobial were used as the source of the dentin discs in these studies. The dentin discs were prepared by attaching the occlusal surfaces of individual teeth, with epoxy resin, to a mounting jig from an Isomet slow-speed diamond saw or by holding the teeth in place by means of a mounting jig. Discs 1.03 + 0.1mm (SEM n=61) thick (Table 1) and approximately 10mm in diameter were sectioned from the crown of the tooth moving from the pulpal to occlusal surface with the Isomet saw. Discs up to and including that containing the last visible remnant of pulp horn (and those showing any evidence of natural decay) were discarded and the next most coronal 1mm disc was used in these experiments. The undecalcified dentin discs were soaked in 0.01M PBS until use.

In order to produce decalcified caries like dentin (Seng et al,

GROWER, ABUEME & SENG
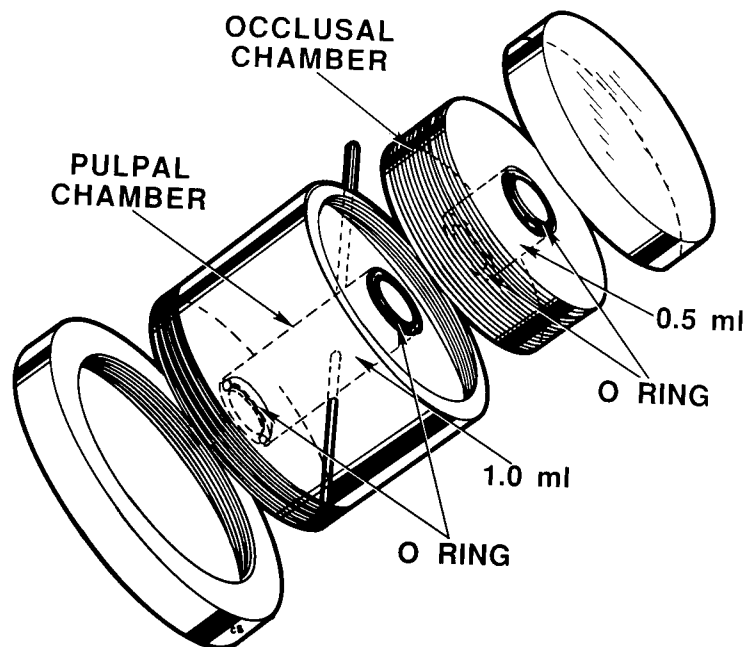
Fig. 1 Schematic drawing of the delrin plastic chamber used as the tooth model. The dentin discs were sealed between the occlusal and pulpal chambers by means of the "O" rings shown.

TABLE 1

CHARACTERISTICS OF DECALCIFIED
DENTIN DISCS USED IN DRUG PENETRATION STUDIES (a)

PROPERTY	MEAN $\pm$ SEM (b)	RANGE
Thickness	1.03 $\pm$ 0.01 mm	0.9 - 1.4 mm
pH 24 hours	6.96 $\pm$ 0.04	6.00 - 7.38
pH 48 hours	6.42 $\pm$ 0.06	5.00 - 7.13

a. The split chambers were incubated at 37° with 0.5 ml of 6% Lactic acid on the occlusal side of the disc and 1 ml of pH 7.6, 0.01M PBS on the pulpal side. The pH was measured after 24 and 48 hrs of acid exposure. b. n=61

GROWER, ABUEME & SENG

1989) the discs were sealed into the delrin chambers with "O" rings with the occlusal surface (with an exposed area of 50.3 mm^2) towards the upper chamber of 0.5ml and the pulpal side (with an exposed area of 19.6 mm^2) facing the lower chamber with a volume of 1ml. The upper occlusal chambers were filled with 0.5ml of 6% Lactic acid and sealed. The lower pulpal chambers were filled with 1.0ml of 0.01M pH 7.6 PBS buffer and the chambers were sealed and incubated 37^0C. pH changes in the PBS solution representing the pulpal tissue were then measured at 24 and 48 hours utilizing a Beckman 701 pH meter and a Fisher micro-electrode. Table 1 shows the results of these preparatory treatments to the dentin.

The surface characteristics of undecalcified, and decalcified dentin discs as well as discs showing natural decay were evaluated using an Amray 1645 Turbo scanning electron microscope.

The penetration of Indomethacin through decalcified and undecalcified dentin was measured by utilizing 4 [^{3}H] Indomethacin labeled agents (Table 2) which were custom prepared by Amersham International using an [^{3}H] exchange process and 4 [^{14}C] Indomethacin labeled agents (Table 3). The ^{14}C Indomethacin compounds were prepared by using 0.3 ml of DMSO to

TABLE 2

COMPOSITION OF [^{3}H] INDOMETHACIN LABELED DELIVERY VEHICLES

Delivery Vehicle		Con. of Unlabeled Indomethacin in Vehicle	Specific Radioacitivity: uCi/mg Indo
1. Petroleum Jelly - Vaseline, Cheseborough Ponds (Hydrophobic)		3 mg/g	2.4 uCi/mg
2. Moisturizing Cream - (Hydrophilic)		3 mg/g	2.4 uCi/mg
3. Diprobase Cream - (Hydrophilic)		3 mg/g	2.4 uCi/mg
Cetomacrogol	2.25%		
Cetostearyl Alcohol	7.2%		
Liquid Paraffin	6.0%		
White Soft Paraffin	15.0%		
Distilled H_2O	7.0%		
4. Unguentum Cream - Merck Co. (Hydrophilic)		3 mg/g	2.4 uCi/mg

GROWER, ABUEME & SENG

dissolve and incorporate 2-14C Indomethacin (Amersham International) and carrier Indomethacin (Sigma Chemical Co.), except for Inteban® which already contained unlabeled Indomethacin, into 10g of each vehicle.

Previous <u>in vitro</u> studies on the permeability of dentin by Greenhill and Pashley (1981) utilized plain Krebs-Ringer's Phosphate buffer as the incubation and elution buffer for tooth chamber systems; however, in studying the dynamics of Indomethacin penetration through dentin a more

TABLE 3

COMPOSITION OF $[^{14}C]$ INDOMETHACIN LABELED DELIVERY VEHICLES

DELIVERY VEHICLES	CON. OF UNLABLED INDOMETHACIN IN VEHICLE	SPECIFIC RADIOACTIVITY; uCi/mg Indo
1. Inteban®Gel (a) (Hydrophilic)	10 mg/g	0.22 & 0.24 uCi/mg
2. White Petrolatum U.S.P. (b) (Hydrophobic)	3.4 mg/g	0.52 uCi/mg
3. Glycerin Compound (b) (Hydrophilic)	3.17 mg/g 3.36 mg/g	0.51 uCi/mg 0.52 uCi/mg
Glycerin 2.5g Polyethylene Glycol 2.0g 3350 Polyethylene Glycol 5.0g 600 Distilled H_2O 1.0g		
4. Polyethylene Glycol Cpd (b) (Hydrophilic)	3.37 mg/g 3.51 mg/g	0.47 uCi/mg 0.52 uCi/mg
Polyethylen Glycol 5.0g 3350 Propylene Glycol 3.12g Distilled H_2O 1.75g		

a. Inteban® Gel, a gel-like ointment, was obtained from Sumitomo Chemical Co., Osaka, Japan. 20 uCi of 2-^{14}C Indomethacin (Amersham International, SA=31mCi/mMol) was added to 10g of the gel to label the compound.
b. 0.30 ml of DMSO was used as a solvent to dissolve 18 or 20 uCi of 2-^{14}C Indomethacin (Amersham International, SA=31mCi/mMol) and 34.4 or 35.8 mg of unlabeled Indomethacin (Sigma Corp.) into each cpd.

GROWER, ABUEME & SENG

physilogic fluid was formulated for use in the pulpal chambers which was more representative of in vivo pulpal fluid. This modification was accomplished by adding 6.5% bovine serum albumin (Sigma Chemical Co.), which approximates the total protein concentration found in vivo in serum (Scientific Tables, 1971), to the 0.01M PBS buffer in order to provide for a colloid effect on the pulpal side of the dentin discs as well as to provide serum albumin for the poorly water soluble Indomethacin to bind to (Yeh, 1985).

Penetration of Indomethacin through the dentin was determined by placing 200mg of each test compound in the occlusal portion of the chamber with a dentin disc sealed in between and with 1ml of pH 7.6 or pH 6.8 0.01M PBS containing 6.5% bovine serum albumin on the pulpal chamber side. The chambers were incubated for 24 hrs at $37^{\circ}C$, the pulpal buffer was removed, and the amount of radioactive Indomethacin which entered the buffer was determined by liquid scintillation counting. The pulpal chambers were re-filled with 1ml of buffer and incubated for an additional 48 hrs to measure further drug release. In some experiments limited equilibrium dialysis studies were done on the delivery vehicles by utilizing two layers of #3 Spectrapore dialysis tubing (Spectrum Medical Industries) with a MW cut off of 3,500 in the chambers rather than dentin discs to measure the diffusional ability of the agents tested. The tubing was pre-soaked in PBS buffer for 12 hours prior to use.

The liquid scintillation data in the form of DPM's of [^{3}H] or [^{14}C] Indomethacin was converted to ug of Indomethacin/ml of buffer by dividing the DPM's observed by the DPM's calculated to be equivelant to 1 ug of Indomethacin for each agent. Statistical analysis of the data was done utilizing one way ANOVA and "t" tests.

RESULTS

Table 1 shows that exposure of the dentin discs to 6% Lactic acid for a period of 48 hours allowed for a gradual acidification of the pulpal buffer as would be seen in the early stages of tissue inflammation. Fig. 2a shows that undecalcified dentin exhibited narrow irregular shaped tubules while exposure of the discs to 6% Lactic acid for 48 hrs produced a decalcification of the surface (Fig. 2b) similar to that exhibited by natural caries (Fig. 2c).

Both [^{3}H] and [^{14}C] labeled Indomethacin delivery vehicles were tested to ensure that the drug penetration effects were due to the actual transport of Indomethacin and not due to the methods used to prepare the radioactively labeled compounds. Table 4 shows the results obtained for the transport properties of 4 custom synthesized [^{3}H] labeled Indomethacin compounds. The 4 [3H] Indomethacin labeled compounds all showed penetration through the decalcified and undecalcified dentin; however none of them showed any significant differences between groups at 24 or 48 hours. Although there appeared to be a tendency for an increase in drug levels present in the pulpal buffer from 24 hrs to 48 hrs for all
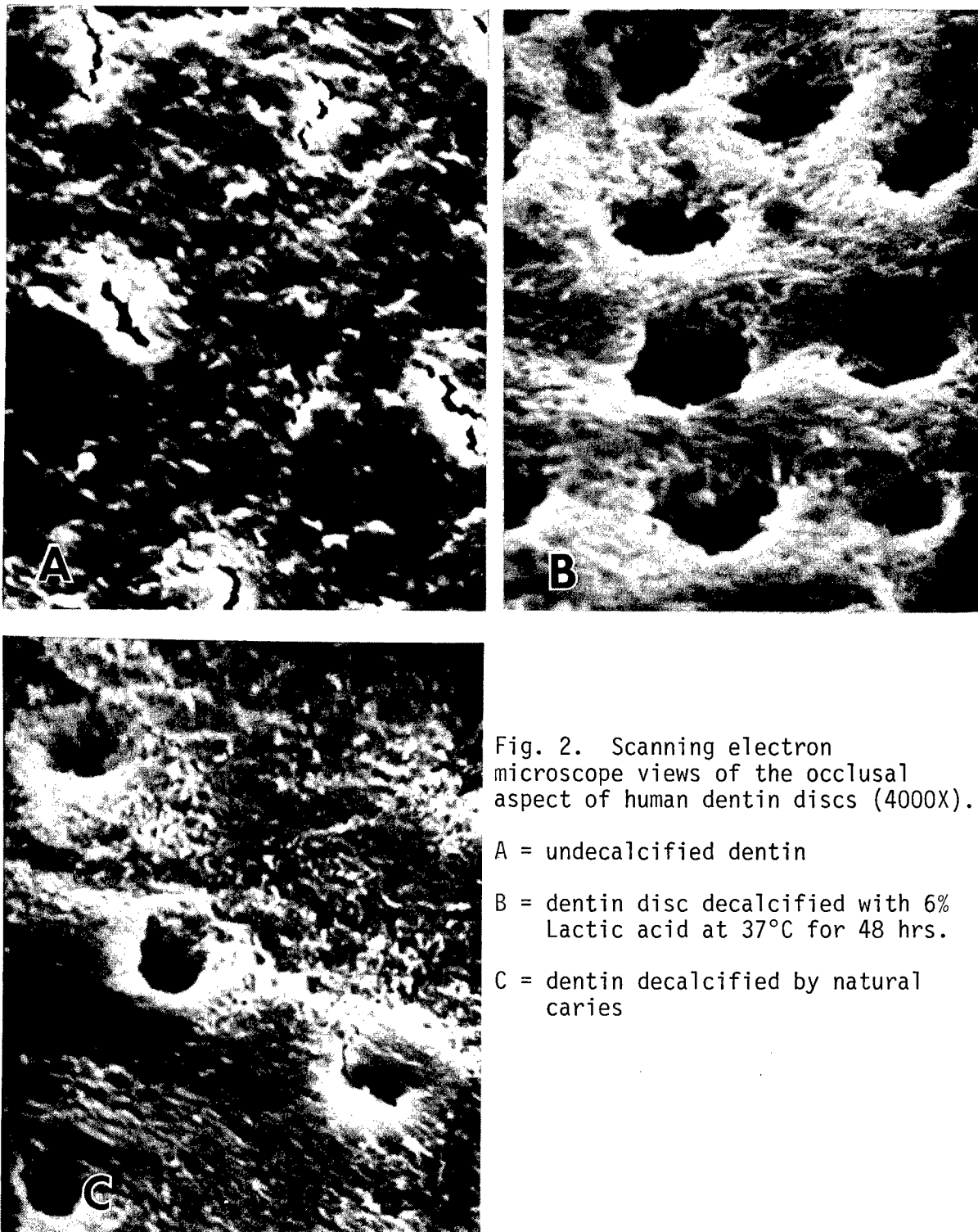
GROWER, ABUEME & SENG

Fig. 2. Scanning electron microscope views of the occlusal aspect of human dentin discs (4000X).

A = undecalcified dentin

B = dentin disc decalcified with 6% Lactic acid at 37°C for 48 hrs.

C = dentin decalcified by natural caries

GROWER, ABUEME & SENG

of the agents, except vasoline, none of these increases within groups was statistically significant. Table 5 shows a summary of the results of Indomethacin delivery using 3 different [^{14}C] Indomethacin labeled agents. Indomethacin appeared to penetrate both decalcified and undecalcified discs as was observed with the [^{3}H] Indomethacin compounds. The delivery vehicle Inteban® which had a level of unlabeled Indomethacin 3 times the other agents (Table 3) carried more than 10-20 times as much Indomethacin through decalcified dentin and 5-10 times as much through undecalcified dentin as the other agents. The statistical analysis in Table 5 shows that this was a significantly higher level than delivered by the other compounds both through decalcified and undecalcified discs. In this experiment the level of Indomethacin delivered by Inteban® at 24 hrs was significantly higher than that observed at 48 hrs; however this difference was not seen in other experiments (Table 7).

Table 6 shows the results of limited equilibrium dialysis experiments done to provide a measure of the water solubility of the delivery agents and their ability to solubilize the Indomethacin incorporated in them. The semi-permeable membranes sealed between the chambers allowed diffusion of

TABLE 4

PENETRATION OF [^{3}H] LABELED INDOMETHACIN
THROUGH HUMAN DENTIN DISCS (a)

COMPOUND	TREATMENT OF DISCS	INDOMETHACIN PENETRATION (b) ug/ml	
		24 hr	48 hr
Unguentum Cream (n=5)	Decalcified Undecalcified	7.03 + 1.95 4.78 ∓ 0.73	8.52 + 2.46 5.89 ∓ 1.13
Moisturizing Cream (n=4)	Decalcified Undecalcified	8.97 + 2.38 9.52 ∓ 2.28	6.22 + 1.74 11.78 ∓ 2.40
Diprobase Cream (n=3)	Decalcified Undecalcified	5.99 + 1.75 4.55 ∓ 1.70	12.03 + 3.52 13.78 ∓ 6.03
Petroleum Jelly (n=2)	Decalcified Undecalcified	6.80 3.90	4.35 2.37

a. Two hundred mg sample of each agent were incubated at 37°C in each chamber with a dentine discs separating it from a chamber containing 1 ml of ph 7.6 buffer with 6.5% bovine serum albumin.
b. Mean ± STD error of the mean.

GROWER, ABUEME & SENG

TABLE 5

PENETRATION OF [^{14}C] LABELED INDOMETHACIN
THROUGH HUMAN DENTIN DISCS - FIRST SERIES

GROUP (a)	DISC TREATMENT	INDOMETHACIN PENETRATION (b) ug/ml	
		24 hr	48 hr
Inteban ®	Decal (n=5)	48.5 + 9.1 (c)	24.6 + 4.1
	Undecal (n=6)	29.9 + 6.5	20.7 + 2.4
Propylene -	Decal (n=3)	4.6 + 0.5	4.7 + 0.7
Glycol CPD	Undecal (n=4)	3.9 + 0.9	4.1 + 0.8
White	Decal (n=4)	2.3 + 0.7	3.0 + 0.2
Petrolatum	Undecal (n=3)	2.6 + 0.6	2.3 + 0.3

a. 200 mg samples of test compounds were applied to each chamber and incubated at 37°C against pH 7.6, (0.01M) PBS containing 6.5% bovine serum albumin.
b. Mean ± SEM.
c. STATISTICAL ANALYSIS:
      ANOVA - Decalcified Discs   (3 groups)
      24 hr   ug/ml, DF between =2, F=16.053, P=0.001
      48 hr   ug/ml, DF between =2, F=16.907, P<0.001

      ANOVA - Undecalcified Discs (3 groups)
      24 hr   ug/ml, DF between =2, F=8.990, P<0.01
      48 hr   ug/ml, DF between =2, F=27.167, P<0.001

t Test   Inteban Decal 24 hrs vs. 48 hrs, P=0.02
t Test   Inteban Undecal 24 hrs vs. 48 hrs, P=0.04
t Tests   Between Inteban®decalcified and Propylene glycol, white petrolatum decalcified at 24 and 48 hrs were significant at P<0.01 to P<0.005
t Tests   Between Inteban®undecalcified and Propylene glycol and white petrolatum undecalcified at 24 and 48 hrs were significant at P<0.01 to P<0.001

GROWER, ABUEME & SENG

the PBS buffer, but not the bovine serum albumin, from the pulpal chambers through the pores of the membrane into the drug containing occlusal chambers. The 3 [^{3}H] Indomethacin labeled compounds (Table 6) showed similar diffusion and binding levels at 24 hours that were about twice that seen when decalcified dentin discs were used (Table 4). The levels of [^{14}C] Indomethacin traversing the membrane with the vehicle Inteban® (Table 6) was similar to that seen penetrating dentin discs (Table 5). Indomethacin diffusion through the dialysis tubing using the vehicle polyethylene glycol (Table 6) was 4 times that through decalcified dentin discs (Table 5) while that of the glycerin compound was 5 times that through undecalcified discs (Table 7) after 24 hours of incubation. At 48 hours the glycerin compound showed even higher levels of Indomethacin transfer. The high delivery rate of the glycerin compound appeared to be due to its very water soluble nature. This resulted in the complete filling of the occlusal portion of the chamber by movement of buffer across the membrane and probably allowed for higher levels of drug transfer seen.

TABLE 6

DIFFUSION OF LABELED INDOMETHACIN THROUGH DIALYSIS TUBING			
COMPOUND	ug INDOMETHACIN IN SAMPLE	INDOMETHACIN DIFFUSION (a) ug/ml	
		24 hr	48 hr
[^{14}C] LABELED			
Inteban®(n=3)	2067	41.4 + 6.0	33.5 ± 2.3
Glycerin CPD (n=4)	641	41.8 ∓ 18.8	92.0 ± 13.1*
Polyethlene Glycol CPD (n=4)	586	21.3 + 4.9**	20.5 ± 8.6
[^{3}H] LABELED			
Unguentum Cream (n=4)	450	15.9 + 1.2	16.8 ± 1.6
Diprobase Cream (n=4)	472	12.9 ∓ 2.8	10.0 ± 3.8
Moisturizing Cream (n=4)	666	13.9 ∓ 4.1	13.2 ± 5.2

a. Mean ± SEM
* t Test   Glycerin CPD 24 hrs vs. 48 hrs $P<0.05$
** t Test   Inteban®vs. Polyethylene glycol at 24 hrs $P<0.05$

In the previous experiments the pH of the pulpal buffer was 7.6; however, the inflamed but still vital pulp would have a more acidic pH

GROWER, ABUEME & SENG

of around 6.8 (Seng & Bayer, 1986). Table 7 shows the results of experiments done to test if there was a difference in Indomethacin delivery according to the pH of the elution buffer in the pulpal chamber. Table 7 shows that at 24 and 48 hours there was no difference in drug delivery with decalcified discs for Inteban®or glycerin compound at pH 7.6 or 6.8. There appeared to be a trend towards less drug penetration through undecalcified discs versus decalcified discs for both compounds tested at pH 6.8 but this was not statistically significant due to the high standard error of the mean observed which reflected the biologic variation of the human dentin discs.

TABLE 7

EFFECT OF pH OF ELUTION BUFFER ON PENETRATION OF [^{14}C] LABELED INDOMETHACIN THROUGH HUMAN DENTIN (a)

TREATMENT OF DENTIN DISCS	pH OF PULPAL CHAMBER BUFFER	INDOMETHACIN PENETRATION (a) ug/ml	
		24 hr	48 hr
INTEBAN ®			
Decalcified	7.6	21.2 + 3.8	15.5 + 0.9
Decalcified	6.8	22.1 + 7.9	22.2 + 3.4
Undecalcified	6.8	10.0 + 1.8	14.0 + 3.1
GLYCERIN CPD			
Decalcified	7.6	7.9 + 1.3	5.9 + 1.2
Decalcified	6.8	6.3 + 3.0	6.3 + 2.0
Undecalcified	6.8	1.2 + 0.2	2.4 + 0.2

a. Mean + STD error of the mean. 4 samples in each group. 200 mg of each agent was used in the occlusal chamber.

DISCUSSION

Indomethacin, with a molecular weight of 358, is an acidic non-steroidal, anti-inflammatory drug. Although practically insoluble in water it is found bound up to 99% to serum albumin (Yeh, 1985) when given orally in therapeutic doses. It has been shown that the dissolution rate of a poorly soluble drug can be increased by dispersing it in a water soluble carrier (Chiou and Riegelman, 1971); therefore the vehicles listed in Tables 2 and 3 were formulated to deliver Indomethacin. One of the methods by which these agents could deliver Indomethacin through the dentin discs to the pulpal buffer would be the creation of an osmotic

GROWER, ABUEME & SENG

pressure difference between the drug and the pulpal chamber which would allow cyclic fluid exchanges between the pulpal and occlusal chambers similar to that seen when hydrogel delivery vehicles are used against skin (Bodde et al, 1989).

Since Indomethacin has pronounced serum albumin binding properties (Yeh 1985; Diana et al, 1989) and dentinal fluid contains proteins similar to those found in plasma (Holdi & Wynn, 1963), the composition of pulpal incubation buffer was formulated to include 6.5% bovine serum albumin which approximates the total protein content of normal blood plasma (Scientific Tables, 1971). This more physiologic solution thus provided a mechanism for transport and stabilization of the poorly water soluble Indomethacin by binding to the serum albumin as well as providing for an osmotic gradient to counteract the water attracting properties of the delivery vehicles. Tables 4 and 5 show that most of the hydrophilic vehicles utilized (Tables 2 and 3) delivered Indomethacin through decalcified dentin to the pulpal chamber at about the same magnitude except for Inteban® which delivered 10-20 times the amount of Indomethacin through decalcified discs than the other compounds listed. Part of this effect may have been due to its higher Indomethacin concentration (3X), however since Indomethacin follows linear pharmacokinetics when given orally (Yeh, 1985) the higher rate of penetration appears to have been due to the nature of the delivery vehicle itself. This positive effect of the penetration enhancers in the vehicle was similar to that seen in underlying muscle after application to skin (Sumitomo Chem. Co., 1981).

The generally similar levels of Indomethacin which were seen penetrating the decalcified and undecalcified discs (Tables 4,5,7), except for Inteban®, may be partly explained by its low molecular weight which would allow it to penetrate the small diameter tubules (Fig. 2a) present even in undecalcified discs as well as the larger tubules (Fig. 2b) in decalcified discs. In addition, the dentin tubules can be regarded as microscopic ion-exchange columns (Pashley, 1988) which may allow for a conduction of Indomethacin through the dentin tubules.

Yeh (1985) showed that.after oral doses of 25 to 75 mg were administered the peak plasma concentration of Indomethacin ranged from 1.2 to 5.8 ug/ml after 1 to 1.5 hours. The results listed in Tables 4,5 and 7 show similar or higher levels of Indomethacin in the pulpal buffer at 24 hours and thus are suggestive that therapeutic levels of Indomethacin may be achievable by local application to thin layers of dentin. These levels of Indomethacin would thus be available to actively reduce the acute inflammatory response present in the pulp due to an acute pulpitis. The higher levels of Indomethacin found in the pulpal buffer when dialysis tubing was used to seperate the incubation buffer from the delivery vehicles, as seen in Table 6 (compared to the results shown in Tables 4,5 and 7) are what would be expected since distance plays a role in diffusion as described by the Fick equation. Within group analysis of all of the compounds tested, except for the glycerin

GROWER, ABUEME & SENG

compound, showed similar levels at 24 and 48 hours suggesting that equilibrium conditions were reached at 24 hours. The higher levels of Indomethacin seen in the pulpal buffer (Table 6) when dialysis tubing was used also provided a measure of the solubility of the delivery agents when water was freely available to hydrate them. In vivo this would be similar to oral administration of the vehicles. The dentin of the tooth, however, does not act the same as a thin, semi-permeable membrane (Pashley, 1989). Thus the in vitro tooth model using actual dentin was used to determine the ability of delivery vehicles to become hydrated and then transfer therapeutic levels of drug through dentin into the pulpal fluid. The use of this system also allowed for observation of the activity of penetration enhancers in the agents, such as seen with Inteban® (Table 5) which allowed transfer of more agent through dentin than would be seen in a simple equilibrium dialysis experiment.

Seng and Bayer (1986) indicated that tissue culture cells exposed to a pH as low as 6.7 to 6.8 could show evidence of an inflammatory response without causing cellular death. The experiments that are summarized in Table 7 were done to confirm that the ability of Indomethacin to penetrate through dentin into pulpal buffer at pH 7.6 would also be exhibited when there was a slight acidic shift as might be encountered when pulpal inflammation was present. The ability to penetrate dentin was not affected by the shift of pH as the data in Table 7 shows. Since Indomethacin is stable in neutral or slightly acidic media, it should also show pharmacologic activity at this pH; however further in vivo tests will have to be done to confirm that these levels are indeed therapeutic.

In this in vitro study serum albumin was used as a component of the pulpal incubation buffer to measure dentin permeability to Indomethacin. Pashley, et al (1982) found in vitro that the plasma protein fibrinogen was one of the most effective agents for producing a reduction in dentin permeability, and they hypothesized that in vivo this might be the agent responsible for acute reductions in permeability and sensitivity seen shortly after cavity preparations. The bovine serum albumin (Cohns Fration V) used in these in vitro studies did not contain fibrinogen, therefore no reduction in dentin permeability due to the pulpal buffers protein content would be expected; however in vivo the presence of fibrinogen may influence the rate of Indomethacin penetration through dentin and the levels of Indomethacin which are obtainable in vivo may vary from those observed in these experiments.

The next phase of these studies will be to investigate the penetration of Indomethacin in an in vivo system to see if the same drug delivery effects are observed that were obtained in the model system and to determine if the Indomethacin so delivered is therapeutically effective in treating pulpal inflammation.

CONCLUSIONS
1. Indomethacin can be formulated into delivery vehicles which will

GROWER, ABUEME & SENG

penetrate both decalcified (caries like) and undecalcified dentin and penetrate into the pulp like chambers.
2. While the amount of Indomethacin penetrating dentin bears a direct relationship to the concentration of drug in the delivery vehicle, the nature of the delivery vehicle itself can also influence its penetration through dentin into the pulp.
3. Indomethacin in suitable delivery vehicles holds promise for in vivo testing as a locally applied agent for treating acute pulpal inflammation and will be investigated in further studies.

REFERENCES

1. Alvarez, O.M., Levendorf, K.D., Smerbeck, R.V., Mertz, P., and Eaglestein, W. Effects of Topically Applied Steroidal and Non-Steroidal Anit-Inflammatory Drugs on Stimulated Repair and Regeneration. Fed. Proc. 43:13, 1984.
2. Bodde, H.E., Van Aaten, E.A.C., and Junginger, H.E. Hydrogel Patches for Transdermal Drug Delivery; In Vivo Water Exchange and Skin Compatability. J. Pharm. Pharmacol. 41:152-155, 1989.
3. Brune, K., Glatt, M., and Graf, P. Mechanism of Action of Anti-Inflammatory Drugs. Gen. Pharmacol. 7:27, 1976.
4. Cassidy, S.E. Why Preventive Dentistry? J. Am. Soc. Prevent. Dent. 7:1970.
5. Chiou, W.L., and Riegelman, S. Pharmaceutical Applications of Solid Dispersion Systems. J. Pharm. Sci. 60:1281-1302, 1971.
6. Diana, F.J., Vernoich, K., and Kapoor, A.L. Binding of Non-Steroidal Anti-Inflammatory Agents and Their Effects on Binding of Racemic Warfarin and Its Enantiomers to Human Serum Albumin. J. Pharm. Sci. 78:195-199, 1989.
7. Ferriera, S., and Vane, J.R. New Aspects of the Mode of Action of Non-Steroidal Anti-Inflammatory Drugs. Ann. Rev. of Pharm. 14:57, 1974.
8. Greenhill, J.D., and Pashley, D.H. The Effects of Desensitizing Agents on the Hydraulic Conductance of Human Dentin in Vitro. J. Dent. Res. 60:686-698, 1981.
9. Haldi, J., and Wynn, W. Protein Fractions of the Blood Plasma and Dental Pulp Fluid of the Dog. J. Dent. Res. 42:1217-1221, 1963.
10. Huskisson, E.C. Classification of Anti-Rheumatic Drugs. In: Anti-Rheumatic Drugs Clin. Pharm. and Ther. Series, Vol 3., ed. E.C. Huskisson, Sussex, UK, 1983.
11. Ingle, J.I., Glick, D., and Schaeffer, L.D. Differential Diagnosis and Treatment of Oral and Perioral Pain. In: Endodontics, 2nd Ed. eds. Ingle and Beveredge. Lea and Febiger, Philadelphia, PA 1976.
12. Ogiso, T., Ito, Y., Iwaki, M. and Atago, H. A Pharmacokinetic Model for the Percutaneous Absorption of Indomethacin and the Prediction of Drug Disposition Kinetics. J. Pharm. Sci. 78:319-323, 1989.
13. Pashley, D.H., Nelson, R., and Kepler, E.E. The Effect of Plasma and

GROWER, ABUEME & SENG

Salivary Constituents on Dentin Permeability. J. Dent. Res. 61:978-981, 1982.

14. Schwartz, C., and Mandel, H. The Selective Inhibition of Microbial RNA Synthesis by Salicylates. Biochem. Pharmacol. 21:771, 1972.

15. Scientific Tables, 7th Edition. Ed. Diem, K., and Lentner, C. pp. 579-580, Ciba-Geigy Limited, Basle, Switzerland, 1971.

16. Seltzer, S., and Bender, I.B. The Dental Pulp, 3rd Ed. J.P. Lippincott, NY, 1984.

17. Seng, G.F., and Bayer, B.M., Inhibition of Amino Acid Transport by NSAID: A Model for Predicting Relative Therapeutic Potency. JPET 237:496-503, 1986.

18. Seng, G.F., Palmer, D.E., and Grower, M.F. Development of a Model System to Measure Drug Penetration Through Decayed Dentin into Inflammed Pulps. J. Dent. Res. 68 (Special Issue/Abstracts) 414, 1989.

19. Simon, J. Pathology. In: Pathways of the Pulp, eds. Cohen and Burns, C.V. Mosby, St. Louis, MO, 1980.

20. Sumitomo Chemical Company LTD. Product Information on Inteban. Osaka, Japan, 1981.

21. Trowbridge, H.O. Changing Concepts in Endodontic Therapy. JADA 110:476, 1985.

22. Yeh, K.C. Pharmacokinetics Overview of Indomethacin and Sustained-Release Indomethacin. AM J. Med., 79 (Suppl 4C) 3-12, 1985.

Supercomputer Simulation of Supersonic Viscous Flow
over Pointed, Spherical, and Flat Tipped Shell

Mr. Bernard J. Guidos* and Mr. Paul Weinacht
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

I. INTRODUCTION

Interest in predicting nosetip bluntness effects on projectile aerodynamics at the US Army Ballistic Research Laboratory (BRL) stems from the fact that almost all Army projectiles are blunt. Designers recognize the fact that nosetip bluntness alters the total drag of shell.[1] Computational and experimental research has shown that nosetip bluntness alters the Magnus characteristics of spinning shell.[2,3] The effects are manifested through detailed flow elements like the three-dimensional boundary layer and bow shock entropy wake. With the computational power of the current generation of supercomputers, the flow can be simulated through large-scale Navier-Stokes computations. An accurate predictive capability will give designers a new tool for determining when the aerodynamic effects of nosetip bluntness are important.

A computational study was undertaken using thin-layer Navier-Stokes techniques to simulate wind tunnel tests done at Princeton University Gas Dynamics Laboratory.[4,5,6,7] The objective was to compare the computational results with the wind tunnel measurements and assess the accuracy of the CFD simulation. Supercomputer resources not previously available were used to conduct the simulations. The wind tunnel model is a tangent ogive-cylinder with a pointed, spherical, or flat nosetip. Comparisons were made of surface pressure, turbulent boundary layer profiles, and nosetip flow structure, and the results are reported herein.

II. WIND TUNNEL CONFIGURATIONS & FLOW CONDITIONS

The computations simulate wind tunnel tests conducted at the Princeton University Gas Dynamics Laboratory 20 cm by 20 cm Mach 3, high Reynolds number, blowdown tunnel. [4,5,6,7] The models, instrumentation, techniques, and data acquisition

and reduction are discussed in detail in the listed references. A brief description is given here to provide a background for the ensuing discussions.

The wind tunnel model and nosetip configurations are shown in Figure 1. The pointed model consists of a 3 caliber (cal) tangent ogive nose and a 5 cal cylinder. The reference diameter (i.e. that of the cylinder) is 4.95 cm (1.95 inches). Two blunt nosetips were included to screw into the basic configuration in place of the pointed nosetip. Both blunt nosetips were machined as a truncation of the pointed nosetip. The pointed nosetip is designated P; the spherical nosetip is designated R3; the flat nosetip is designated F3. The bluntness ratio, defined as the ratio of fuze tip diameter to reference diameter, is 25%. In the case of a spherical nosetip, the fuze tip diameter is defined as twice the radial distance from the axis to the juncture of the sphere and ogive. For a flat nosetip, the fuze tip diameter is defined as twice the radial distance from the axis to the juncture of the flat face and ogive.
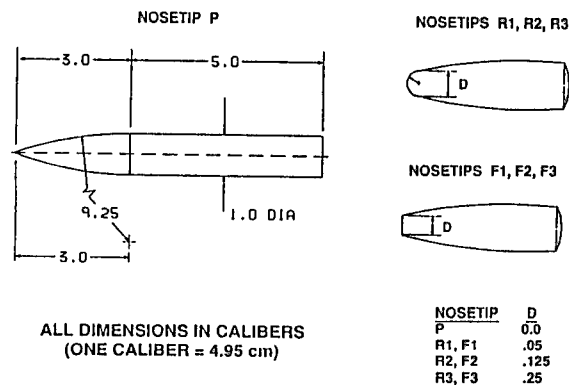
Figure 1. Model and Nosetip Configurations

The model was supported at the base by an axial sting and held fixed relative to the free stream. Wall pressure distributions, pitot pressure surveys, and shadowgraphs were obtained. Longitudinal (streamwise) velocity profiles were generated from the pitot surveys which were taken at three axial locations on the cylinder. The shadowgraphs provided a means for locating the bow shock, the boundary layer transition zone, and, in the case of a flat nosetip, the imbedded recompression shock.

Conditions for the computations are taken as the average over numerous wind tunnel runs. The Mach number is 2.95, with angles of attack of 0° and 2.9°. The nominal stagnation temperature is 260K, the nominal Reynolds number is 3.14×10^6 based on model diameter, and the stagnation pressure is 6.89×10^5 N/m (6.8 atm). The wall temperature is taken as adiabatic and the flow is assumed to be steady. Boundary layer transition occurred in the experiment without the use of a tripping device, and varied widely between the different nosetips. The computational simulation of transition is discussed in a subsequent section.

120

III. COMPUTATIONAL APPROACH

The major advantage of this study compared to the study of Reference 2 was the availability of Cray X-MP and Cray 2 computers. These resources were especially valuable in the case of the flat nosetip, where substantial grid resolution was implemented in order to capture the streamwise flow separation which was expected to occur.

The computational methodology consists of two separate techniques applied in succession. An unsteady Navier-Stokes (UNS) technique is used to compute the flow in the vicinity of each blunt nosetip. These solutions provide initial conditions for a parabolized Navier-Stokes (PNS) technique, which computes the flow by integrating, or marching, in the streamwise direction. The use of these two techniques is an efficient way to model the flow because the unsteady (time-iterative) technique, which is considerably more computationally intensive, is applied over a relatively small region near the blunt nosetip. The parabolized (space-marching) technique, which processes only two physical planes of data simultaneously, is applied over the remaining downstream region, excluding the base.

Computational modeling of the flow over the blunted nosetips was accomplished using the thin-layer UNS approach first reported by Pulliam and Steger.[8] This technique integrates the transformed, time-dependent, thin-layer, Navier-Stokes equations in strong conservation law form, given as:

$$\frac{\partial \hat{q}}{\partial \tau} + \frac{\partial \hat{E}}{\partial \xi} + \frac{\partial \hat{F}}{\partial \eta} + \frac{\partial \hat{G}}{\partial \zeta} = \frac{1}{\hat{R}e} \frac{\partial \hat{S}}{\partial \zeta} \tag{1}$$

Equation (1) represents conservation of mass, momentum, and energy in the transformed coordinate directions, ξ, η, and ζ (Figure 2), for large Reynolds number flow. The vector $\hat{q}$ is the transformed vector of dependent variables. The vectors $\hat{E}, \hat{F}$, and $\hat{G}$ are the transformed inviscid flux vectors. The vector $\hat{S}$ is the vector of viscous terms that result from the thin-layer approximation. The Reynolds number is $\hat{R}e$. The dependent variables are $\rho, \rho u, \rho v, \rho w$, and e: density, momentum components in the x, y and z directions, and total energy per unit volume, respectively. The governing equations are obtained at each grid point using an approximately-factored, implicit, delta-form, finite-difference algorithm.[9] Second order central differencing is used in the three coordinate directions. Fourth order explicit and second order implicit smoothing terms are added to suppress high frequency oscillations in the solution.[10] The solution at each time step requires a series of block-tridiagonal matrix inversions (sweeps) in each of the transformed coordinate directions. Complete details of the governing equations and numerical method are available in the referenced literature.

Computational modeling of the flow over the ogive and cylinder portion (and the pointed nosetip) was accomplished using the thin-layer PNS technique first reported by Schiff and Steger.[11] This technique spatially integrates the transformed, steady, thin-layer Navier-Stokes equations in strong conservation law form, given as:

$$\frac{\partial \hat{E}_s}{\partial \xi} + \frac{\partial \hat{F}}{\partial \eta} + \frac{\partial \hat{G}}{\partial \zeta} = \frac{1}{\hat{Re}} \frac{\partial \hat{S}}{\partial \zeta} \qquad (2)$$

Equation (2) represents the thin-layer approximation to steady-state conservation of mass, momentum, and energy in the three transformed coordinate directions.

Figure 2. Coordinate Transformation

$\hat{E}_s$, $\hat{F}$, and $\hat{G}$ are the inviscid flux vectors. $\hat{E}_s$ is a modified flux vector resulting from the subsonic sublayer approximation.[11]

The dependent variables $\rho, \rho u, \rho v, \rho w$, and e are obtained at each grid point using the approximately factored, implicit, delta form, finite-difference algorithm.[9] Second order central differencing is used in the circumferential and radial directions, and first order one-sided differencing is used in the marching direction. Fourth order explicit smoothing is added to suppress high frequency oscillations in the solution. The solution is advanced downstream by numerically integrating in the direction parallel to the projectile axis. Each spatial (marching) step requires a series of block tridiagonal matrix inversions (sweeps) in the circumferential and radial directions. Complete details of the governing equations and numerical method are available in the referenced literature.

IV. NUMERICAL GRIDS

Figure 3 shows the computational grid for nosetip R3 (converged solution, $\alpha = 0°$). The free-stream flow is left to right, and the outer boundary represents the converged, fitted bow shock. The x coordinate is measured from the virtual origin, defined as the intersection of a fully extended ogive and the model axis. The ogive extension, with lines of constant $\xi - \eta$ becoming radially oriented downstream of the nosetip, facilitates coupling with the PNS code. For the axisymmetric flow cases, a single circumferential plane is used. For the angle of attack cases, where bilateral symmetry exists, 21 circumferential planes were distributed from $\phi = -10°$ to $\phi = 190°$ at 10°

increments.

Figure 4 shows the computational grid for nosetip F3 (converged solution, $\alpha = 0°$). This wrap-around grid consists of straight line segments between the body and shock. At the corner of the flat face, the constant $\xi - \eta$ line intersects the body at a 45° angle with respect to the projectile axis. The flat nosetip geometry is arrived at during the flow field solution procedure by truncating a small portion of the (initially) spherical nosetip, re-generating the grid and bow shock, and allowing the solution to relax completely (about 500 time steps). The procedure is repeated several times until the final body shape is obtained.
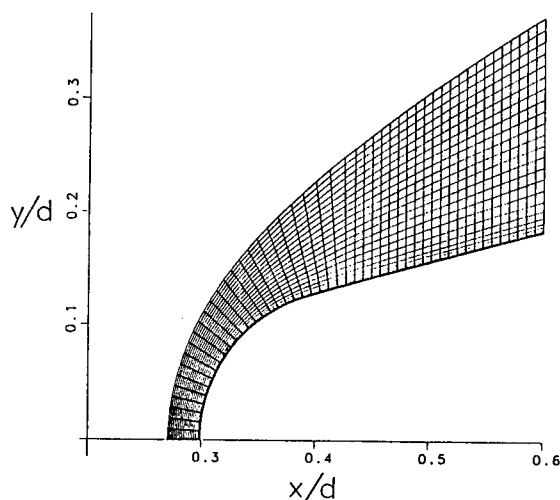
Figure 3. Spherical Nosetip Grid
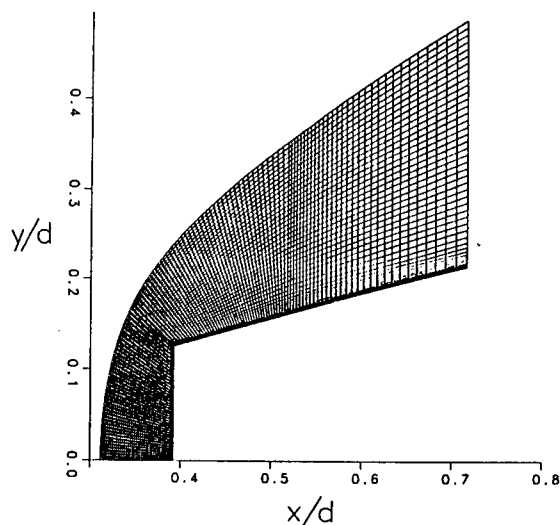
Figure 4. Flat Nosetip Grid

The PNS computations were performed using cylindrical grids. In the radial direction, 45 grid points were used with exponential clustering near the body. For solutions having bilateral symmetry, circumferential planes were distributed from $\phi = 0°$ to $\phi = 180°$ at 10° increments. Some additional cases were run using 5° increments, but the flow properties of interest remained unaffected. For axisymmetric solutions, only three circumferential planes were used at $\phi = 0°$, $\phi = 90°$, and $\phi = 180°$ in order to employ a cylindrical coordinate formulation[12] of the PNS equations. The grid density in the marching direction varied from 300 to 500 marching steps over a length of approximately 8 cals.

V. BOUNDARY LAYER TRANSITION MODEL

The pitot surveys in the experiment were taken at three different axial locations on the cylinder, where the flow was turbulent. Boundary layer transition can be seen in the shadowgraphs to vary as a function of nosetip geometry, angle of attack, and roll angle. Overall, transition began approximately 0.5-2.5 cal from the virtual origin and extended approximately 0.25-0.5 cal downstream. The sensitivity of the numerical results to location and length of transition was investigated using the shadowgraphs as a guide. Laminar flow conditions were used in all of the UNS computations. Transition was simulated in the PNS computations in the following manner:

The effective viscosity, usually defined as

$$\mu = \mu_l + \mu_t \tag{3}$$

is re-defined as

$$\mu = \mu_l + \kappa_t \mu_t \tag{4}$$

where μ_l is the laminar (molecular) viscosity, μ_t is the turbulent viscosity obtained from the Baldwin-Lomax model,[13] and κ_t is the transition intermittency factor. This factor is identically zero for laminar flow and identically one for fully turbulent flow.

From probability considerations for a turbulence spot production function, Dhawan and Narasimha[14] derived a distribution of transition intermittency that agreed with a survey of wind tunnel measurements over a flat plate. The distribution took the form

$$\kappa_t = 1 - \exp(-A(x - x_t)^2) \quad \text{for} \quad x > x_t \tag{5}$$

where A is a positive constant that represents the rate of transition and x_t is the location where transition begins. The the location and length of transition are not prescribed a priori, but are instead empirically based. A similar distribution which is more easily applied to parametric study is the polynomial

$$\kappa_t = 10 \left(\frac{x - x_1}{x_2 - x_1} \right)^3 - 15 \left(\frac{x - x_1}{x_2 - x_1} \right)^4 + 6 \left(\frac{x - x_1}{x_2 - x_1} \right)^5 \quad \text{for} \quad x_1 < x < x_2 \tag{6}$$

where x_1 and x_2 mark the beginning and end of transition. The properties of this polynomial are (1) it varies from zero to one in the domain of interest, (2) it has zero slope at the endpoints, and (3) it is symmetric about the midpoint. Equation (6) was used in the PNS computations to simulate transition. Effects due to surface curvature, pressure gradient, or nosetip geometry itself, are not accounted for in the model. Variations with respect to roll angle due to model incidence are also ignored.

VI. COMPARISON WITH EXPERIMENT

1. Nosetip Flow Structure

The bow and recompression shock locations were measured in this study using transparencies of the shadowgraphs overlayed onto a fine Cartesian mesh. Severe optical distortion from the large density gradients at the nosetip prohibited direct measurement relative to the vertex of the actual nosetip. Reference marks were not made on the tunnel window or on the model itself, compounding the difficulty in obtaining quantitative data. Instead, the apparent bow shock vertex was used as the reference point for the readings. The uncertainty in the measured distance between any two points in the shadowgraphs was found to be approximately ±0.002 cals. The uncertainty in the absolute radial location of the measured points was found to be approximately ±0.002 cals. The uncertainty in the absolute axial location of the measured points varied up to ±0.006 cals.

Figure 5 shows the computed Mach contours for nosetip R3, $\alpha = 0°$, along with the computed and measured bow shock locations. The computed bow shock location agrees with experiment to within the accuracy of the measurement (which is discussed subsequently). The computed Mach contours illustrate the smooth expansion that occurs around the nosetip for this case. Such smoothness in the flow field is a computational asset because, compared to a flat nosetip, less grid resolution is required.
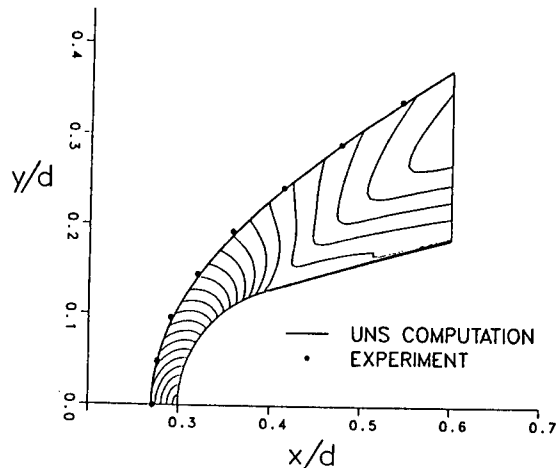
Figure 5. Bow Shock and Mach Contours, R3

Figures 6 and 7 show the computed Mach contours for nosetip F3, $\alpha = 2.9°$, along with the computed and measured shock structures. The strong expansion at the corner is visible in the computations, but the recompression shock is less obvious. The computed imbedded shocks are instead visualized in the plot as thickened solid lines. These lines represent surfaces where the velocity component in the direction of the pressure gradient is equal to the local sound speed, and the Mach number in that direction is decreasing.[15] The shock relations are not presumed to be satisfied, as evidenced by the presence of the small, disconnected patches near the body.
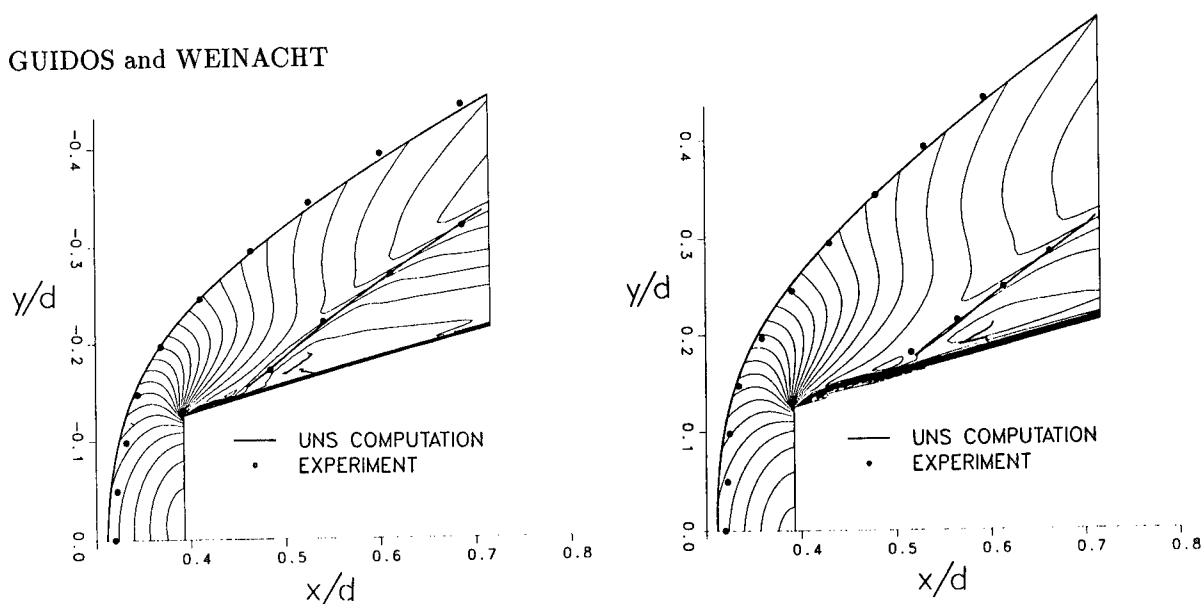
Figure 6. Shocks and Mach Contours, F3, Wind   **Figure 7. Shocks and Mach Contours, F3, Lee**

The comparison of the computed and measured shock patterns in Figures 5 through 7 is within the accuracy of the measurement in each case. Because of the limits in measurement accuracy previously discussed, the positions of the bow and recompression shocks relative to each other may be more indicative of agreement than the absolute comparisons. The trends with respect to incidence are well-predicted along the pitchplane, with the bow shock moving closer to the body on the wind side and farther from the body on the lee side. The recompression shock can be clearly seen to move upstream on the wind side and downstream on the lee side. In Figure 6, the slight bend in the recompression shock is apparent in both computation and experiment. The comparisons lend credence to the previously mentioned method of locating imbedded shocks in a computational flow field solution.

Whether flow separation occurs around the shoulder of the flat nosetip is difficult, if not impossible, to determine from the shadowgraphs. Figure 8 shows a close up view of the computed velocity vectors for nosetip F3, $\alpha = 0°$. The computation shows that the flow separates at the corner and reattaches approximately one tenth of a cal downstream. A secondary separation region is computed to exist within the main region. Figures 8 and 9 illustrate the need for a high density grid near the corner to adequately resolve the separation region. The supercomputer memory is essential for performing computations with such a large number of grid points.
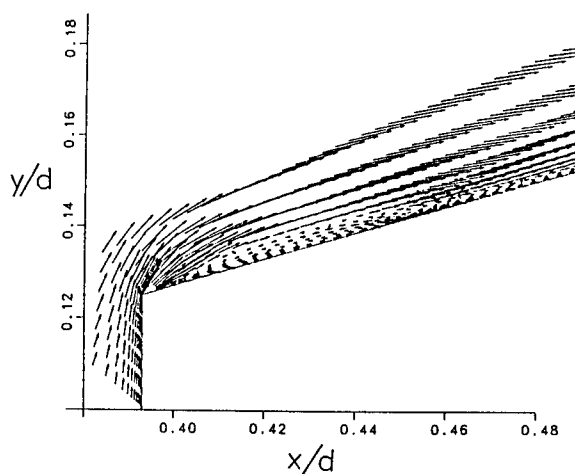
Figure 8. Velocity Vectors, F3
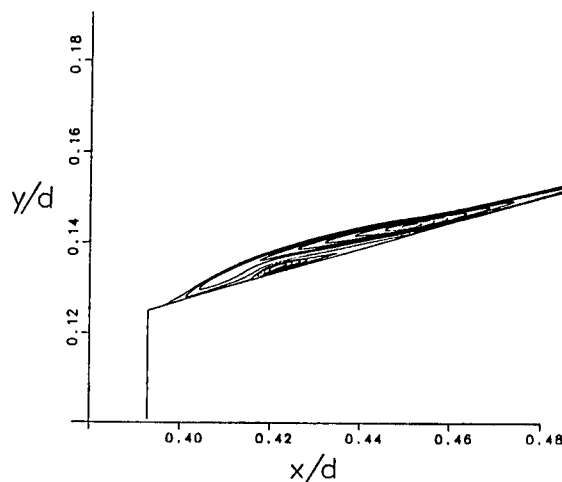
Figure 9. Particle Traces, F3

2. Surface Pressure

The accuracy of the surface pressure measurements is reported in Reference 6 to be typically ±1.0% of the free stream static pressure. Larger variations in the data, due to changing stagnation conditions over the duration of the experiments, are indicated by Reference 6. For nosetip F3, $\alpha = 0°$, the stagnation point pressure varies in different runs from p/p_∞ =11.38 to p/p_∞ =11.72, about ±3%. Downstream, different transducers were used, and the largest variation indicated from the tables also occurs for nosetip F3, $\alpha = 0°$. At the sixth pressure tap, at $x/d \approx 0.48$, the measured value of pressure varies in different runs from p/p_∞ =1.452 to p/p_∞ =1.603, about ±5%. These two examples are important indicators of experimental accuracy in the sense that nominal free-stream conditions were used in the computations.

Figure 10 shows the comparison of the computed and measured surface pressure for nosetip R3, $\alpha = 2.9°$. A smooth variation of pressure along the body is indicated. The flow does not overexpand. The agreement between computation and experiment is within the experimental accuracy, and the influence of small angle of attack is consistent.

Figures 11 and 12 show the comparison of the computed and measured surface pressure for nosetip F3, $\alpha = 2.9°$. Near the stagnation point, the pressure agrees to within the experimental accuracy. The third pressure tap, located on the flat face of the nosetip, shows that the locally subsonic flow is already expanding before it reaches the

corner. Whereas the computation shows a moderate disagreement with the measurement at this tap location, the severe expansion appears to be qualitatively captured in the numerical solution. A closer examination of the grid and the Mach contours reveals that much of this expansion is captured near the body surface between two adjacent grid points, indicating that agreement might be improved by adding, or clustering, grid points even closer to the corner. The tail end of the computed expansion shows a slight overshoot before recovering to a pressure level more comparable to that of the first tap downstream of the corner.
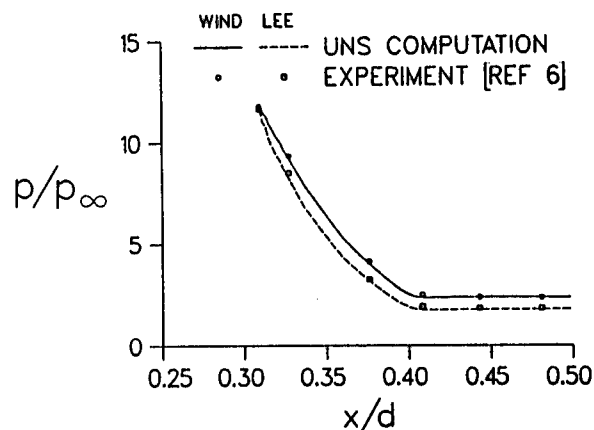
Figure 10. Surface Pressure Comparison, R3
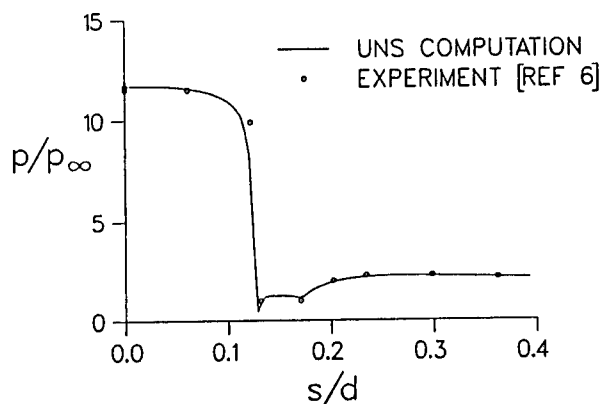
Figure 11. Surface Pressure, F3, Wind
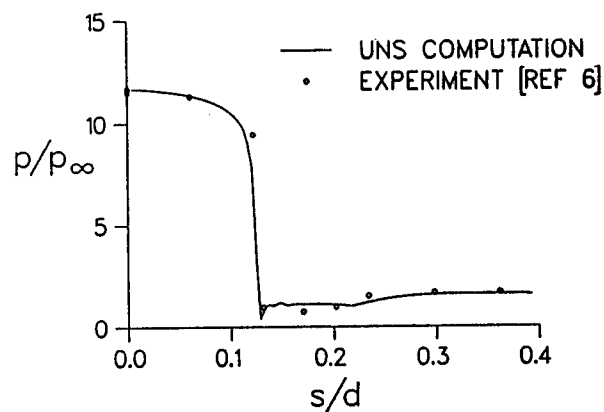
Figure 12. Surface Pressure, F3, Lee

Downstream of the corner, the flow overexpands and recompresses in both the computation and the experiment. The computed pressures in the overexpanded region compare with the fifth pressure tap to within the measurement accuracy for $\alpha = 2.9°$, wind side. On the lee side, the measured pressure is slightly lower than the computed pressure. The locations of the recompression correspond well on the wind side. On the lee side, the computed recompression at the surface occurs slightly farther downstream than in the experiment. In all cases, however, the computed surface pressure recovery

levels agree with the experiment to within the measurement accuracy.

The pressure distribution and flow structure at the nosetip are important indicators of the localized effects of bluntness. Farther downstream, as reported in References 5-7, the surface pressure shows a variation of only ±1.5% between the seven different nosetips. A more sensitive indicator of bluntness must be utilized for purposes of computational validation, i.e the velocity profiles.

3. Longitudinal Velocity Profiles

The comparisons of computed and measured velocity profiles presented here inherently contain uncertainties. For the measured profiles, these include effects of instrument calibration shifts, model misalignment, probe bending, assumptions used for reducing pitot pressure surveys to velocity profiles (for example, that of constant static pressure across the boundary layer), and variations in stagnation conditions both during and between runs. In the comparisons presented here, nosetip P is included as a reference for discerning the effects of bluntness on the profiles. A comparison of the blunt nosetip profiles relative to the pointed nosetip profiles is an important consideration in determining how well the bluntness effects are being modeled computationally.

Figures 13 and 14 show the velocity profile comparisons for nosetips P and R3, $\alpha = 2.9°$. Throughout most of the boundary layer, the absolute agreement between
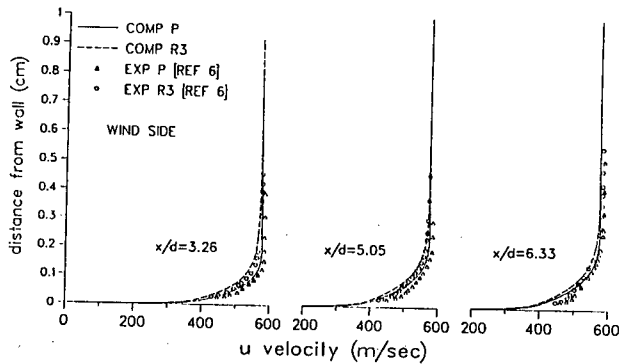
Figure 13. Velocity Profiles, P and R3, Wind
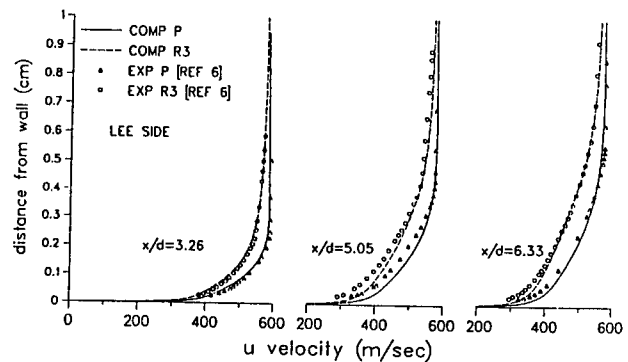
Figure 14. Velocity Profiles, P and R3, Lee

computation and experiment is within 5%. The exception is at the second station on the lee side, where the agreement is within only about 8-9%. In all cases, however,

the comparison worsens within about 0.05 cm from the wall, but with no clear trend. The comparisons of the R3 profiles relative to the P profiles show favorable agreement. There is close agreement with the recovery that the R3 profile exhibits over the three axial stations. On the wind side, both computation and experiment show the velocity deficit of the R3 profile relative to the P profile to be minimal. On the lee side, both show the boundary layer edge of the spherical nosetip profiles to extend outward beyond the edge of the pointed nosetip profiles. There is also agreement in the outward migration of the merging of the profiles as x/d increases. The slight bulge in the the measured wind side profiles at x/d=6.33, 0.05 cm from the wall, occurred for all nosetips and may be an interference effect.

Figures 15 and 16 show the velocity profile comparisons for nosetips P and F3, $\alpha = 2.9°$. Throughout most of the boundary layer, the absolute agreement between computation and experiment is within about 3%. The exception, again, is at the second station on the lee side, nosetip P, where the agreement worsens. And again, in all cases, the comparison worsens within 0.05 cm from the wall, but with no clear trend. Similar to the R3 profiles, the comparisons of the F3 profiles relative to the P profiles show favorable agreement. Both computation and experiment show the velocity deficit for the flat nosetip to be greater than for the spherical nosetip. On the wind side, there is favorable agreement with the velocity deficit in the F3 profiles, in the recovery of the F3 profiles relative to the P profiles over the three axial stations, and in the inward migration of the radial location where the profiles merge. On the lee side, the effects of blunting are seen to be most drastic. Both computation and experiment show the lack of recovery of the F3 profile relative to the P profile.
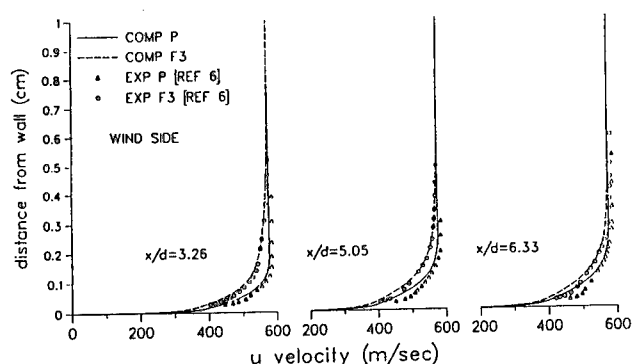
Figure 15. Velocity Profiles, P and F3, Wind
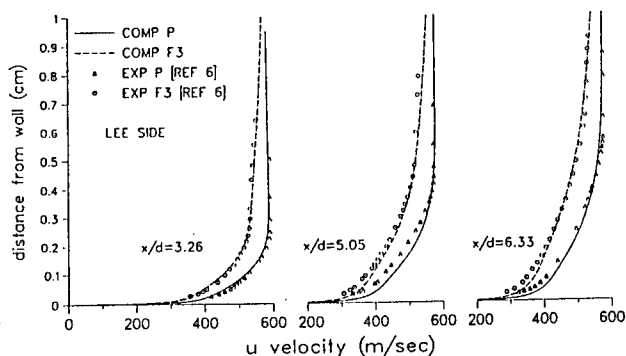
Figure 16. Velocity Profiles, P and F3, Lee

VII. SUMMARY

A thin-layer Navier-Stokes computational study has been made to simulate the effects of nosetip bluntness on a body of revolution at Mach 2.95 . The objective was to evaluate the accuracy of the computational approach through comparison with experimental data. Two thin-layer Navier-Stokes codes were used in conjunction to compute the flow over the model (excluding the base region). The model consisted of a pointed, spherical, or flat nosetip with up to 25% bluntness, followed by a 3 caliber tangent-ogive and a 5 caliber cylinder. The pointed nosetip provided the reference case for discerning the effects of bluntness. The computational grid for the blunt nosetips was generated using a single-zone, algebraic formulation that wraps the grid around the body to conform with the surface. Boundary layer transition was simulated in the computations based on shadowgraphs which were taken during the experimental study.

For the spherical nosetip case, the comparisons between computation and experiment were mostly within the measurement accuracy for bow shock location and surface pressure distribution. The comparisons of turbulent velocity profiles on the cylinder agreed within about 1% to 5%, except close to the wall. The agreement was substantially better when comparing the profiles relative to the pointed nosetip case. Solutions were obtained on a Cray X-MP/48 using 1 to 3 hours of CPU time and using less than a half million words of memory.

For the flat nosetip case, the comparisons between computation and experiment were mostly within the measurement accuracy for bow shock location, recompression shock location, and surface pressure distribution. The computed surface pressure showed some disagreement close to the corner where a rapid expansion occurs, but still showed strong qualitative agreement. The comparisons of turbulent velocity profiles on the cylinder agreed mostly within 1% to 3%, except close to the wall. The computation showed the flow to separate around the corner of the flat nosetip, a flow detail which is difficult, if not impossible, to discern from the shadowgraphs. The flow separation is believed to strongly influence the downstream profiles. The complexity of the flat nosetip flow field is exemplified by the fact that 40 hours of CPU time and 2.5 million words were used on a Cray-2 for the $\alpha = 2.9°$ case. The large amount of computer usage was necessary to obtain an accurate benchmark solution.

The study has demonstrated that the flow details of pointed, spherical, and flat tipped shell can be predicted to a high degree of accuracy. Through comparison with experiment, it has been shown that a single-zone, wrap-around, algebraic grid can be employed to model the laminar flow over a nearly square corner. The computations serve as a benchmark for future CFD efforts in the modeling of supersonic flow over blunted projectiles. The accurate prediction of Magnus characteristics of spinning,

flat-tipped shell may now be in hand. One of the remaining challenges is to reduce the computational intensiveness of the problem while ensuring that the accuracy of the flow details is not compromised.

References

1. McCoy, R.L., "'McDrag'- A Computer Program for Estimating the Drag Coefficients of Projectiles," U.S. Army Ballistic Research Laboratory/ARRADCOM, Aberdeen Proving Ground, MD, ARBRL-TR-02293, February 1981. (AD A09810)

2. Sturek, W.B., Mylin, D.C., Guidos, B.J., and Nietubicz, C.J., "Navier-Stokes Computational Study of the Influence of Shell Geometry on the Magnus Effect at Supersonic Speeds," U.S. Army Ballistic Research Laboratory/ARRADCOM, Aberdeen Proving Ground, MD, ARBRL-TR-02501, June 1983. (AD A130630)

3. Kayser, L.D., "Effects of Small Nose Bluntness on Static Stability and Magnus Characteristics of a Projectile Shape at Mach 0.91 and 3.03," U.S. Army Ballistic Research Laboratory/AMC, Aberdeen Proving Ground, MD, BRL-MR-3535, July 1986. (AD A171734)

4. Dolling, D.S. and Bogdonoff, S.M., "Supersonic, High Reynolds Number Flow over a Tangent Ogive Cylinder at Small Angles of Attack: An Experimental Study and Comparison with Theory," Princeton University Mechanical and Aerospace Engineering Department, Princeton, New Jersey, Report No. 1421-MAE, January 1979.

5. Gray, W.K., Dolling, D.S., and Bogdonoff, S.M., "An Experimental Investigation of Tip Bluntness Effects on the Turbulent Compressible Boundary Layer on an Axisymmetric Body," Princeton University Mechanical and Aerospace Engineering Department, Princeton, New Jersey, Report No. 1530-MAE, May 1981.

6. Dolling, D.S. and Gray, W.K., "Compilation of Wall Pressure and Turbulent Boundary Layer Data for Supersonic, High Reynolds Number Flow over a Blunted Tangent Ogive Cylinder at Small Angles of Attack," Princeton University Mechanical and Aerospace Engineering Department, Princeton, New Jersey, Report No. 1585-MAE, September 1982.

7. Dolling, D.S. and Gray, W.K., "Experimental Study of Supersonic Turbulent Flow on a Blunted Axisymmetric Body," AIAA Journal, Volume 24, No.5, May 1986.

8. Pulliam, T.H. and Steger, J.L., "Implicit Finite-Difference Simulations of Three-Dimensional Compressible Flow," AIAA Journal, Vol. 18, No. 2, February 1980, pp. 159-167.

9. Beam, R. and Warming, R.F., "An Implicit Factored Scheme for the Compressible Navier-Stokes Equations," AIAA Paper No. 78-0257, AIAA 16th Aerospace

Sciences Meeting, January 1978.

10. Pulliam, T.H. and Steger, J.L., "Recent Improvements in Efficiency, Accuracy, and Convergence for Implicit Approximate Factorization Algorithms," AIAA Paper No. 79-0130, AIAA 17th Aerospace Sciences Meeting, January 1979.

11. Schiff, L.B. and Steger, J.L., "Numerical Simulation of Steady Supersonic Viscous Flow," AIAA Journal, Vol. 18, No. 12, December 1980, pp. 1421-1430.

12. Rai, M.M. and Chausee, D.S., "New Implicit Boundary Procedures: Theory and Applications," AIAA Paper No. 83-0123, AIAA 21st Aerospace Sciences Meeting, January 1983.

13. Baldwin, B.S. and Lomax, H., "Thin-Layer Approximation and Algebraic Model for Separated Turbulent Flows," AIAA Paper No. 78-0257, AIAA 16th Aerospace Sciences Meeting, January 1978.

14. Dhawan, S. and Narasimha R., "Some Properties of Boundary Layer Flow During the Transition from Laminar to Turbulent Motion," Journal of Fluid Mechanics, Vol. 3, part 4, pp 418-436, 1958.

15. Buning, P.G., private communication concerning PLOT3D plotting package, Version 3.3, NASA Ames Research Center, Moffett Field, CA, 1987.

Characterization of the Catalytic Site for G Agent Hydrolysis Using Electron Paramagnetic Resonance (U)

*John W. Halliday, Mr.
John E. Walker, Mr.
Soldier Science Directorate
U.S. Army Natick RD&E Center
Natick, MA 01760-5020

K.S. Rajan, Dr.
IIT Research Institute
10 West 35th Street
Chicago, IL 60616

INTRODUCTION

The exploration of several new technologies will provide the technical basis to augment the protection of the soldier, both individually and collectively, from the diverse hazards of operation in a chemical agent contaminated environment. These technologies seek to complement and supplement the current reliance on the barrier properties and sorptive capabilities of materials used in protective device construction. Prominent among these new systems are those based on reactive and catalytic materials - especially those employing metal complexes. In the research and, to a more limited extent, the developmental phases, the reactivity of such systems towards chemical agents and their reaction surrogates is studied in solution. To be of practical use, however, these systems must be immobilized by, for example, inclusion in a bio- or synthetic-polymer matrix. The challenge is to relate the reactivities and associated mechanisms between these two dissimilar physical conditions.

Metal-ligand systems are an appropriate model for study in approaching the structure surrounding the metal center in a polymer matrix. Divalent copper chelate systems, the subject of the present study, are accessible to several types of spectroscopic investigation - electron paramagnetic resonance (EPR) techniques being especially useful. The copper center is an *in situ* paramagnetic probe sensitive to various types of relevant perturbations. Divalent copper, Cu(II), is a $3d^9$ system with one unpaired electron in the valence shell. Cu(II) is a hexacoordinated system with six ligand molecules arranged around the copper center in a basically octahedral symmetry. The five-fold degeneracy of the 3d electron orbital manifold (holding a maximum of 10 electrons) is lifted by the field of these six ligands. Two orbital subsets result, t_{2g} and e_g; the unpaired electron is in the latter set. The energy states of Cu(II) and how these states are perturbed by the various ligands can be approached using the constructs of crystal field, ligand field and molecular orbital theory.

In most practical cases the symmetry of the coordinated Cu(II) is reduced to axial symmetry by distortion along a C_3 axis perpendicular to a plane containing the Cu center and four ligand electron donor sets. The minimal effect of the apically coordinated ligands (usually H_2O) on the spectroscopy of Cu(II)[1] results in the familiar square planar representation for the coordination configuration.

EPR spectra are quantitatively characterized (Figure 1) by the spectroscopic factor, g, and the hyperfine coupling constant, A. For axial symmetry, the spectroscopic g factor and the hyperfine coupling constant, A, are anisotropic with $g_x = g_y = g_\perp$, $A_x = A_y = A_\perp$ and $g_z = g_{||}$, $A_z = A_{||}$. Thus, the EPR spectrum experimentally observed (Figure 1) is characterized by a *perpendicular* portion and a *parallel* portion. The paramagnetism of Cu(II) systems is best observed at low temperatures, e.g. < 160°K; thus, the EPR spectra obtained in this study are powder-type spectra, i.e., the paramagnetic centers are randomly oriented (spatially) and are relatively immobile.

The line widths expected for powder spectra are further broadened by the unresolved contributions of the ^{63}Cu and ^{65}Cu isotopes in their natural abundance (69.2% and 30.8%, respectively). Due to the broad lines and small $A_\perp$ values, the perpendicular portion of the spectrum is unresolved. The occurrence of extra artifact lines at high field [2] makes even the determination of $g_\perp$ problematical. For these reasons, interpretation of the EPR spectra is based on the parallel portion.

Figure 1. EPR spectrum of Cu(II) aquo complex in an ethylene glycol glass at 150°K

The interaction between the ligand and the Cu(II) is manifested in both the g values and the hyperfine coupling constants. Deviation of the g value from the free electron value (g=2.0023) is effected by a spin orbital contribution from the ligand donor atoms. Thus, the nature of the ligand affects the g value, i.e., the "position" of the spectrum on the magnetic field axis. Information bearing on the type and "strength" of the Cu(II) - ligand bond is evidenced in the magnitude of the hyperfine coupling constant, the interaction of the unpaired electron with a nucleus having a net moment. In the present case, only the hyperfine interaction of the unpaired electron with the ^{63}Cu and ^{65}Cu nuclei (both nuclear spin 3/2 systems) is observed. The broadness of the lines precludes observation of superhyperfine interactions from, for example, ^{14}N. Although the magnetogyric ratios of the two copper isotopes differ by about 7%, separate contributions from each isotope could not be resolved and the observed lines are thus averages.

The ligands included in this study are N,N,N',N'-tetramethyl ethylenediamine (TMEN), glucosamine (GA), and 1-methoxy glucosamine (MGA). The structures of these ligands are:

$(CH_3)_2N$ — $N(CH_3)_2$

H_2C—CH_2

N,N,N',N'-tetramethyl
ethylenediamine
TMEN

glucosamine
GA

1-methoxy glucosamine
MGA

EXPERIMENTAL

EPR spectra were taken using a Bruker Model ER 200 D-SPC Electron Spin Resonance Spectrometer equipped with an ESP 1600 data system, an ER041MR microwave bridge and a BVT-1000 Variable Temperature controller. An ER 4103 TM X-Band Microwave Resonator, a flat cylindrical cavity resonating in the TM $_{110}$ mode with a nominal 9.8 GHz center frequency, was used. The magnet was equipped with 10-inch diameter pole faces with specially contoured caps to correct field inhomogeneities resulting from the presence of 1 cm diameter axial holes. Scans of ±750 gauss centered at 3000 gauss (with a field resolution of 0.18 gauss) were normally made at a scan rate of 1.1 gauss/second (a nominal 20-minute scan). Longer scan times were occasionally required to achieve adequate signal/noise ratios. The normal 100 kHz modulation frequency was used with a modulation amplitude of 2 gauss (occasionally 5 gauss was used for extremely weak signals). Power saturation effects were not observed; consequently, a 20 mW excitation level was used throughout.

EPR measurements were made at two temperatures: 77°K and 150°K. The lower temperature was achieved using a special liquid nitrogen (LN$_2$) dewar. For the higher one, the BVT-1000 Variable Temperature system, which generates a stream of cold nitrogen gas by controlled boil-off from a reservoir of LN$_2$, was used. This cold gas stream is conveyed by a series of transfer dewars to an open ended sample-holding dewar situated in the microwave cavity. Temperatures in the range 145-150°K were the lowest that could be achieved and reliably maintained using this technique.

The EPR investigation of the binary Cu(II) chelate systems tetramethylethylenediamine (TMEN), glucosamine (GA), and 1-methoxyglucosamine (MGA) and the ternary systems Cu(II):TMEN:GA and Cu(II):TMEN:MGA was carried out in both a glassy matrix at 150°K and in a polycrystalline aqueous matrix at 77°K. Gradual freezing of sample solution maintained in a piece of thin quartz tubing (ca. 3mm internal diameter) using cold vapors from LN$_2$ produced a poly-microcrystalline sample. Rapid freezing of the dilute aqueous solutions by submersion in LN$_2$ resulted in highly fractured samples (due to the wide discrepancies in the expansion coefficients) which were useless for EPR analysis. Samples in a glassy matrix were produced from a 1:1 dilution of the aqueous sample with ethylene glycol. In this case, rapid submersion in LN$_2$ produced a glassy, transparent sample of near optical quality.

The EPR spectra obtained in the glassy system were in all cases dramatically better resolved (Figure 2) than corresponding spectra in a polycrystalline matrix.
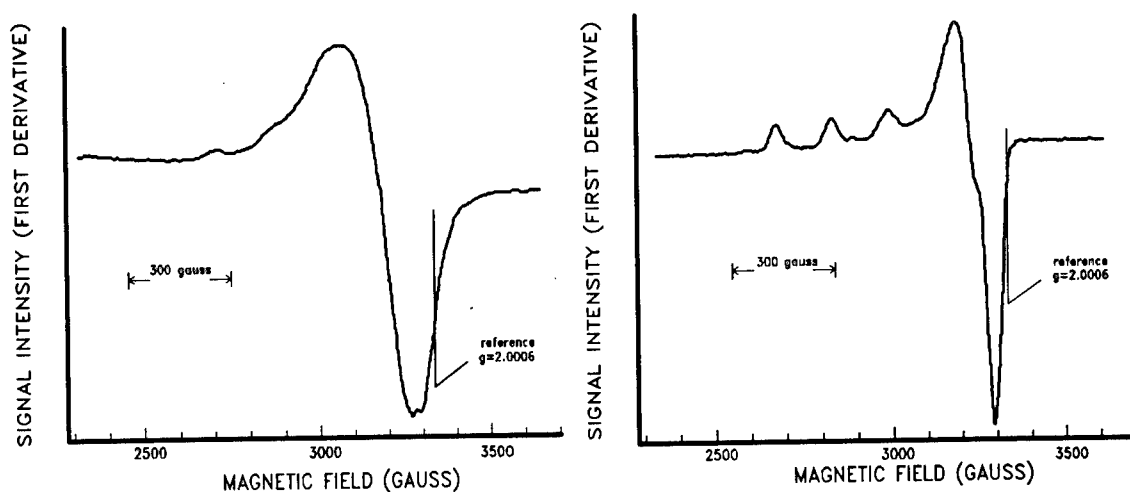
Figure 2. EPR spectra for Cu:TMEN:MGA, 1:1:1 at pH=6.5 in a polycrystalline matrix at 77°K (left) and in a 1:1 ethylene glycol glass matrix at 150°K (right).

An internal g-value standard was used in all measurements - a color center produced in the quartz tube by exposure to ionizing radiation (high energy (10 MeV) electrons or ^{60}Co γ-rays). The positive going lobe was used as the reference point. The g-value for this feature (g=2.0006) was determined by direct comparison (Figure 3) with the commonly used g-value standard 2,2-diphenyl-1-picrylhydrazyl free radical (DPPH), g=2.0037 [3].
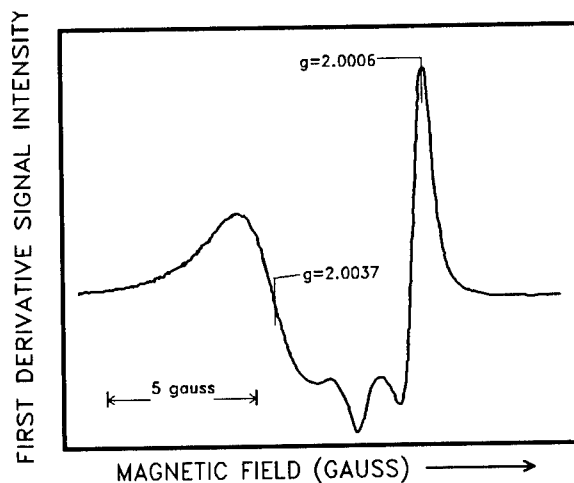
Figure 3. EPR spectrum of the color center in γ-irradiated quartz. The correction in g-value between DPPH (g=2.0037) and the positive lobe of the color center (g=2.0006) is illustrated. The DPPH contribution to the spectrum is at lower field and the asymmetric quartz color center part is at higher field.

The position of hyperfine features relative to this internal standard for the parallel portion of the experimental spectra was determined using spectrum manipulation features of the Bruker ESP 1600 data system. Features were assigned to groups or patterns based on comparison and contrast with selected basis spectra, e.g., $Cu^{+2}(aq)_6$ and TMEN 1:1 (pH=6.0 and pH=7.2).

The relative positioning of these patterns is shown in Figure 4. The value of $A_{\parallel}$ for each group was based on the two lowest field hyperfine features. Splittings between higher field features (when resolution permitted) was used as confirmatory data. From these measurements the center of the parallel region of the spectrum for that group was located and a simple algorithm calculated the g-value, $g_{\parallel}$, for this position. Figure 1 illustrates, graphically, the positioning of $g_{\parallel}$ for an uncomplicated spectrum in which only one species is present. As indicated previously, little useful quantitative information can be gleaned from the perpendicular (high field) part of the spectrum.

Solutions of the variously coordinated Cu(II) species were prepared using copper salts with the naturally occurring isotopic distribution: ^{63}Cu (69.2%) and ^{65}Cu (30.8%). Copper and ligands were mixed in the appropriate molar ratios, i.e., 1:1 or 1:2 for the binary chelates and 1:1:1 for the ternary, and the pH adjusted to the desired value.
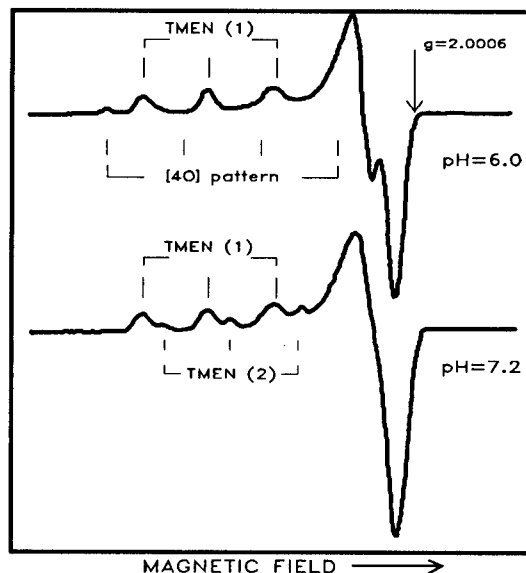
Figure 4. EPR spectra of TMEN 1:1 pH=6.0 & pH=7.2 in an ethylene glycol glass at 150°K. Base patterns TMEN (1) and TMEN (2) as well as the superposition of the Cu(II) aquo ([4O] pattern) line positions are illustrated.

RESULTS

The basic agreement of averaged $A_{\parallel}$ and $g_{\parallel}$ values derived from the EPR spectra obtained for both the polycrystalline matrix at 77°K and the 1:1 ethylene glycol glass matrix at 150°K confirm the absence of significant matrix effects. Consequently, the better resolved and more revealing spectra obtained for the glassy matrix at 150°K will form the basis for the ensuing analysis.

The several combinations of pH and metal:ligand molar ratios resulted in 10 binary systems and six ternary systems for EPR investigation. The results are summarized in Table 1. The ratio $g_{\parallel}/A_{\parallel}$ is cited as an indicator of the amount of tetragonal distortion[2]. Complexes with ratios > 135 cm are believed to exhibit such distortion. In our case (Table 1 and Figure 5) these are limited to the aquo-type species.

Table 1. Values of the EPR parameters $g_{\parallel}$ and $A_{\parallel}$ derived for the discernible species in binary and ternary Cu chelate systems involving TMEN, GA, and MGA at various pHs and molar ratios. $A_{\parallel}$ is in mK (10^{-3} cm^{-1}); the ratio $g_{\parallel} / A_{\parallel}$ has units of cm.

Ligands	Ratio	pH	Peak	$A_{\parallel}$ (mK)	$g_{\parallel}$	$g_{\parallel}/A_{\parallel}$ (cm)
TMEN	1:1	6.0	-	17.9	2.298	128
TMEN	1:1	7.2	1	18.3	2.297	126
			2	19.0	2.247	118
GA	1:1	6.0	-	14.1	2.420	171
GA	1:2	6.5	1	18.7	2.308	123
			2	14.4	2.418	168
GA	1:2	7.0	1	19.3	2.303	119
			2	17.6	2.409	137
MGA	1:1	6.0	-	14.3	2.418	169
MGA	1:1	6.5	-	14.3	2.418	169
MGA	1:2	6.5	1	19.4	2.298	119
			2	15.0	2.410	161
MGA	1:2	7.2	-	17.6	2.299	131
MGA	1:2	8.0	-	20.8	2.255	109
TMEN-GA	1:1:1	6.5	1	18.2	2.298	126
			2	19.1	2.246	117
TMEN-GA	1:1:1	7.2	1	19.0	2.294	121
			2	19.7	2.241	114
TMEN-GA	1:1:1	8.0	1	18.7	2.302	123
			2	19.8	2.241	113
TMEN-MGA	1:1:1	6.5	-	18.1	2.297	127
TMEN-MGA	1:1:1	7.5	1	17.6	2.298	131
			2	19.6	2.247	114
TMEN-MGA	1:1:1	8.0	1	17.7	2.296	130
			2	19.1	2.250	118

Ideally, the EPR results would be used to determine parameters appropriate to crystal field or ligand field models and thereby provide a quantitative measure of the Cu-ligand interaction. Powder-type data, such as our spectra provide, do not have sufficient resolution or detail to support this approach[4]. The most productive treatment is that of J. Peisach and W.E. Blumberg[1]. Their empirical approach based on an extensive survey of copper complexes uses plots of $g_{\parallel}$ versus $A_{\parallel}$ to relate the independent changes in $g_{\parallel}$ and $A_{\parallel}$ to changes in the nature of the complexation.

A summary of all the data, both binary and ternary systems, in a Peisach/Blumberg style plot (Figure 5) of $g_{\|}$ versus $A_{\|}$ reveals clustering in three g value regions - $g_{\|}$ = 2.24-2.25, $g_{\|}$ = 2.30, and $g_{\|}$ = 2.40-2.42. The Cu(II)aquo complex is in this last cluster. This clustering would, at first glance, indicate three basic types of complexation. A more detailed analysis of each group as well as the significance of the intergroup differences requires several interpretative guidelines. A shift in the g factor towards a lower value indicates increased contribution from a more "electron rich" donor set, such as nitrogen, to the Cu-ligand interaction. The effect of decreasing charge within a given donor set is to shift the g value downward and to increase the hyperfine coupling energy - an upward and to the left offset on the plot.
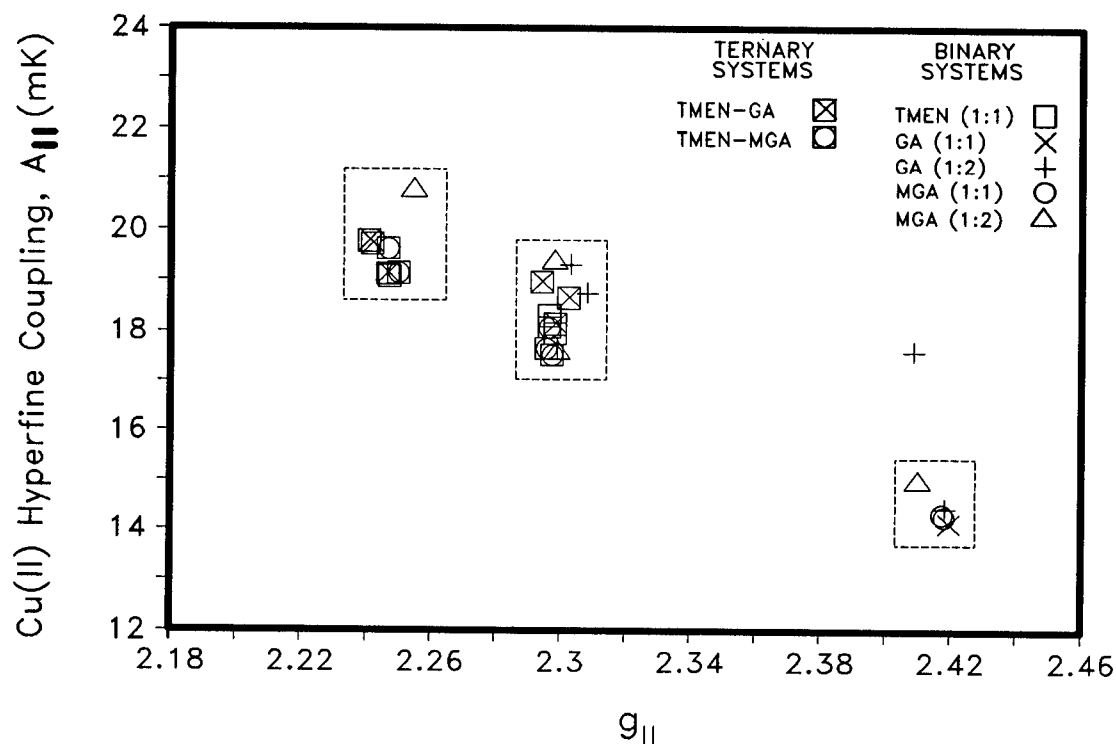
Figure 5. Plot of $g_{\|}$ versus $A_{\|}$ for all the binary and ternary systems studied. The hyperfine coupling constant is in mK (10^{-3} cm^{-1}).

The Binary Systems. The results for the binary systems Cu:TMEN (1:1), Cu:GA (1:1), Cu:GA (1:2), Cu:MGA (1:1), and Cu:MGA (1:2) are summarized as a Peisach/Blumberg style plot in Figure 7. With the pH variants, there are 10 such systems approximately 50 percent of whose EPR spectra indicate more than a single species present. The ligands are all bidentate; that is, each ligand molecule has two atoms that can serve as potential electron donors to the copper-ligand interaction - two nitrogens for TMEN and a nitrogen and oxygen for GA and MGA. The inclusion of MGA as a ligand provides the opportunity to distinguish between hydroxyls in the C-1 and in the C-3, C-4 positions and their involvement in the ligation. Free of the constraints imposed by consideration of pH and metal to ligand ratio, there are a multitude of bonding possibilities depending on which ligand molecules can be involved in the bonding and which donor atoms can contribute.

The scheme in Figure 6 identifies structures for some of the potential complexes [5,6,7] as the pH of the system increases. While the pH and molar ratio regimes used restrict the number of possibilities for a given chelate system, a distribution of species exists under most experimental conditions. These different species are reflected in the EPR spectra as distinctive patterns.

Figure 6. Some chelate structures for binary Cu:GA, Cu:MGA, and Cu:TMEN systems. The designation in the brackets ([1N]) is the donor set.

The cluster in the $g_{||}$ =2.42, $A_{||}$ =14-16 mK region for the pH=6.0 and 6.5 is due to the aquo-type complexes. These complexes are observed for several systems - Cu:GA(1:1), Cu:GA(1:2), Cu:MGA(1:1), and Cu:MGA(1:2). For the Cu:GA(1:1) and Cu:MGA(1:1) these are the sole species observed. Species distribution data based on potentiometric titration techniques [5,6,7,8] indicate that from 50-90% of the metal is present in an unchelated form, i.e., the aquo complex in this pH range. Even for the TMEN system at pH 6.0, H_2O is able to compete for the Cu(II) and the aquo complex is observed (Figure 4).

For TMEN (1:1) at pH=7.2, only nonaquo species are observed - two species, in fact. The pattern for one species is the same as observed at pH=6.0. At this lower pH, the species is predominantly due to the donor set [1N,1N'], i.e., coordination with two TMEN molecules through the [NH2] donor set. At the higher pH (7.2) contributions from a mixture of species, including the donor set [2N,1N'], are possible. The EPR parameters do not allow us to distinguish among

142

these species. The second species at pH=7.2, however, is consistent with the [2N,2N'] donor set. The offset towards both lower g-value and higher hyperfine coupling energy indicates involvement of four nitrogens.
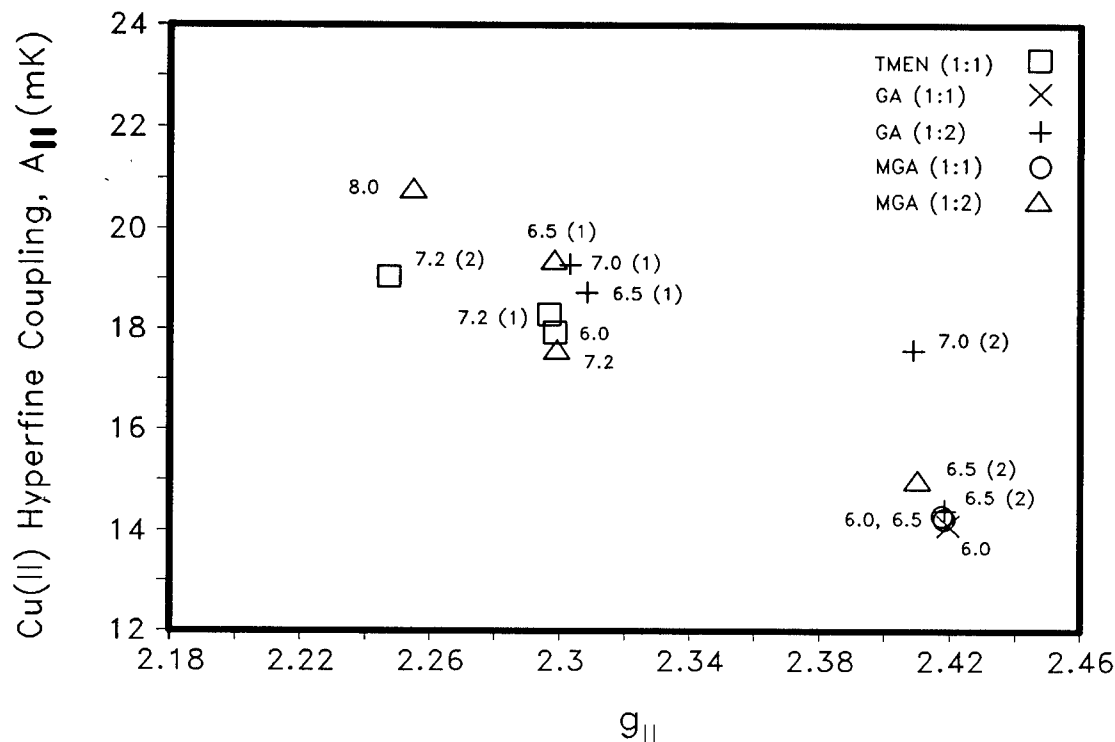
Figure 7. Plot of $g_{||}$ versus $A_{||}$ for the binary systems. The hyperfine coupling constant is in mK (10^{-3} cm^{-1}). Each entry has been annotated with its pH and EPR pattern (in parentheses) as designated in Table 1.

For GA(1:2) at the higher pHs (6.5 -7.0) there is a second species present. The $g_{||}$ =2.308, $A_{||}$ =18.7 mK (pH=6.5) and $g_{||}$ =2.303, $A_{||}$ =19.3 mK (pH=7.0) are consistent with the EPR parameters reported by Micera et al.[6] for the CuL_2 species. The [NH_2] donor sets are from different GA molecules. The species distribution data differ on the exact amounts of these species present. The other possibility is the $CuLH_{-1}$ species - a [1N,1O] donor set. Whether the slightly higher $g_{||}$ and $A_{||}$ values (compared to those for TMEN) supports this interpretation is problematic. The parameters from the Peisach/Blumberg plots for these two donor sets[1] predict little difference in either the value for $g_{||}$ or the interaction energy. While CuL_2 is the major contributor in this general pH range[7], CuL_2H_{-2} may also present at the higher end of the range (pH 7.0). Our results do not, however, show the shift in $g_{||}$ predicted for this species. A small contribution from this species may account for the slightly higher $A_{||}$ at pH 7.0. It should be noted and emphasized that two species with different $g_{||}$ values but present in disproportionate amounts may be manifested in the EPR spectrum as slightly broader, but unresolved lines.

143

Results for the MGA systems qualitatively parallel those for GA. At the lower pHs (6.0 and 6.5), the MGA(1:1) spectra are attributable solely to the aquo species. For the MGA(1:2) system at pH 6.5, the aquo is one contributing species; the other is probably CuL. The spectra of the MGA(1:2) series are considerably less intense and resolved than for any of the other systems. This situation is a caveat with regard to the precision and accuracy of the $g_{\parallel}$ and $A_{\parallel}$ determinations.

At the higher pHs (7.2 and 8.0), only one species can be clearly discerned. The resolution is, however, so poor that pursuing an interpretation based on the $g_{\parallel}$ and $A_{\parallel}$ shifts is not merited.

The ternary systems. The results for the two ternary systems Cu:TMEN:GA (1:1:1) and Cu:TMEN:MGA (1:1:1) are summarized in the Peisach/Blumberg style plot in Figure 8. The data cluster in two fairly narrow regions around $g_{\parallel} = 2.3$ and $g_{\parallel} = 2.25$. The $A_{\parallel}$ values are also fairly narrowly clustered.
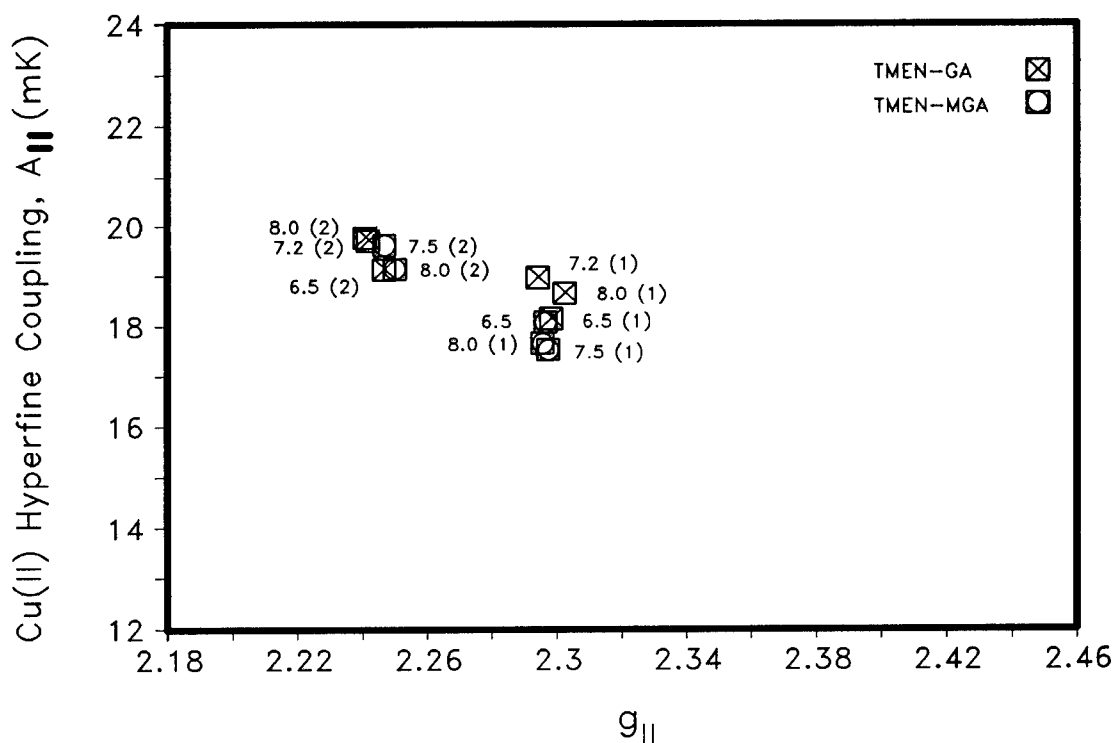
Figure 8. Plot of $g_{\parallel}$ versus $A_{\parallel}$ for the ternary systems. The hyperfine coupling constant is in mK (10^{-3} cm^{-1}). Each entry has been annotated with its pH and EPR pattern (in parentheses) as designated in Table 1.

Both ternary systems are characterized by two species; or, to be more precise, by two collections of species with comparable EPR parameters. The relative amount of each species is pH dependent. As shown in Figure 9, pattern (1) with $g_{\parallel} = 2.294 - 2.302$ and $A_{\parallel} = 17.6 - 19.0$ mK predominates at low pH. As the pH increases, pattern (2) ($g_{\parallel} = 2.241 - 2.250$, $A_{\parallel} = 19.1 - 19.8$ mK) assumes dominance. The shift in the average $g_{\parallel}$ from the 2.30 to the 2.25

region and the slight increase in the average $A_{\parallel}$ is consistent with a change in donor sets from [2N]/[3N] to [4N]. The involvement of a [2N,2O] donor set from CuL_2H_{-2} is also consistent. Except for the aquo complexes, the contributing species at the lower pHs may, in principle, be any combination of those discussed for the binary systems. The completely coordinated mixed ligand case, i.e., one TMEN molecule ([2N]) and one GA molecule ([1N,1O]), a [3N,1O] donor set, would also be in the ensemble of species at $g_{\parallel} = 2.30$. The effect of increasing pH favors both the involvement of more nitrogen donor sets from TMEN and the aforementioned potential contribution from CuL_2H_{-2}. The ternary system may be, in a simplistic view, a source of nitrogen donor sets. The species distribution data [5,8] (obtained, again, by potentiometric titration methods) indicates a shift from a primarily TMEN-dominated ligation at lower pH to a mixed ligand complex at the higher pHs.
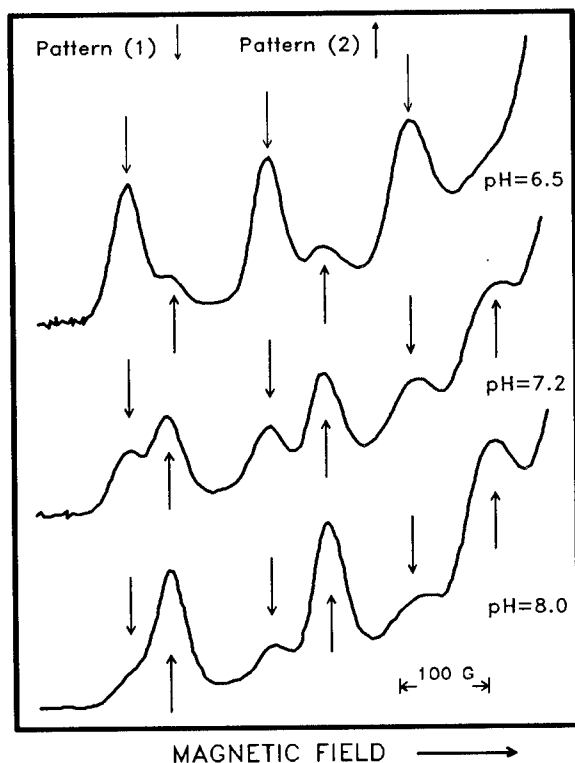
Figure 9. The change in the species distribution with pH for the ternary system Cu:TMEN:GA as reflected in the EPR spectra. This figure shows only the low field ($g_{\parallel}$) part of the spectrum. Pattern (1) is indicated by ↓, pattern (2) by ↑.

The distribution among the two EPR pattern types (based on peak height measurements) for both Cu:TMEN:GA and Cu:TMEN:MGA is given in Table 2.

Table 2. Distribution of species for the Cu:TMEN:GA and Cu:TMEN:MGA ternary systems as derived from the EPR spectra.

pH	TMEN:GA		TMEN:MGA	
	% pattern 1	% pattern 2	% pattern 1	% pattern 2
6.5	80	20	95	5
7.2	40	60	-	-
7.5	-	-	46	54
8.0	18	82	34	66

In contrast to the binary system Cu:MGA(1:2), EPR spectra for the ternary system involving MGA (Cu:TMEN:MGA) were of similar good intensity and quality to those involving GA.

DISCUSSION

The spectroscopic identification and characterization of specific species are relevant to a consideration of the reactivity of Cu(II) chelate systems only to the degree that identification and characterization illuminate our understanding of the factors controlling that reactivity. The hydrolysis/detoxification of phosphonate chemical agents by Cu(II) complexes, even in solution, undoubtedly depends on many factors. One of the authors (K.S. Rajan) has reported [5,8] thermodynamic stability constants for a number of Cu(II) chelates and the rates of hydrolysis of chemical agents by them. The relative hydrolysis rates of GB at pH=7.2 by the systems considered in this study are reported as:

Chelate system	Normalized Hydrolysis Rate
Cu:TMEN:MGA (1:1:1)	1.0
Cu:TMEN (1:1)	0.79
Cu:TMEN:GA (1:1:1)	0.73
Cu:GA (1:1)	0.31
Cu:MGA (1:1)	0.24
Cu:MGA (1:2)	0.22
Cu:GA (1:2)	0.22

Except for TMEN (1:1) there is a marked contrast in rates between the binary and ternary systems. The order of reactivity between the two ternary systems is in the same direction as contribution of the pattern (1) species to the total distribution (Table 2). At pH=7.2 the

Cu:TMEN:MGA pattern (1) contribution is estimated at 57% of the total. This is 42% greater than the pattern (1) content in the Cu:TMEN:GA system (40%). The hydrolysis rates differ by 37%. It would thus seem that the transition from the [2N]/[3N] donor set cluster to the fully coordinated [4N] is a factor. TMEN, it will be recalled (Figure 4), exhibited the partition between patterns (1) and (2) as a function of pH. At pH=7.2 the predominant species was pattern (1). The pattern (1) species in both TMEN and the ternary systems is characterized by an average $g_{||}$ / $A_{||}$ ratio of 126 cm. The average ratio for the pattern (2) species is slightly, but consistently, lower - 115 cm. Thus, there also appears to be a slight difference in distortion between the two species.

The complexation site for Cu(II) in the biopolymer chitosan is characterized[2] by $g_{||}$ =2.244 and $A_{||}$ =18.5 mK ($g_{||}$ / $A_{||}$ =121 cm). These values are very close to our pattern (2) and consistent with a [4N] donor set. That our pattern (1) type species exhibit the enhanced reactivity would indicate the need for an unencumbered coordination site to support the transition state.

Most of the questions associated with the correlation of species (or ensemble of species) with the observed EPR parameters have already been discussed and will recieve no additional discussion. Other spectroscopies can be used to augment and further illuminate the characterization offered here. However, EPR spectroscopy is in a unique position to be able to study the Cu(II) centers in all their divergent physical environments. Thus, the present characterization is a basis with which to compare and contrast the behavior of Cu(II) in these more relevant matrices.

REFERENCES

1. J. Peisach and W. E. Blumberg, Arch. Biochem. & Biophy., 165, 691-708, 1974.

2. S. Schlick, Macromolecules, 19 (1), 192-195, 1986.

3. J. E. Wertz and J. R. Bolton, *Electron Spin Resonance*, 1972, McGraw-Hill, Inc.

4. J. F. Boas, J. R. Pilbrow, and T. D. Smith, *Biological Magnetic Resonance* (ed. L. J. Berliner & J. Reuben), Chapter 7, 1978, Plenum Press.

5. K.S. Rajan, S. Mainer, J.E. Walker, *et al.*, *Proceedings of the 1989 Conference on Chemical Defense Research*, Chemical Research, Development and Engineering Command, in preparation.

6. G. Micera, S. Deiana, A. Dessi, *et al.*, Inorganica Chimica Acta, 107, 45-48, 1985.

7. A. Pusino, D. Droma, P. Decock, *et al.*, Inorganica Chimica Acta, 138, 5-8, 1987.

8. K.S. Rajan, private communication.

Vulnerability Analysis of Protective Structures
Using Probabilistic Methods (U)

Ms. Pamela G. Hayes
U.S. Army Engineer Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, Mississippi 39180-6199

Introduction

The most difficult phase of the design and modification of protective structures and facilities is determining their vulnerability or survivability to the expected threat. In the past, large-scale testing has been used to determine the vulnerability of selected targets. A large number of full-scale tests are required to achieve an accurate measure of vulnerability, and the cost of these tests continues to rise. As an alternative to full-scale testing, computer-modeling of the target and threat is being used for probabilistic determination of the target vulnerability. A computer code, COBRA, has been developed at the U.S. Army Engineer Waterways Experiment Station (WES) for the probabilistic vulnerability analysis of protective structures, personnel, and some equipment. This code is unique in the detail with which targets can be defined and their cumulative damage can be identified. The expected threat is modeled with the errors associated with the particular weapon system, the projectile characteristics, and an attack definition. The vulnerability of the target is then determined based on the probability of damage from direct hit, blast, and fragmentation. After all rounds in an attack have been evaluated the target is rebuilt, and the attack process is repeated until the standard deviation, standard error, and confidence intervals meet the desired limits.

COBRA was developed and used at WES in the late 1970's and early 1980's. Then, in 1987, studies were resumed that utilized the capabilities of the code. Modifications were made to allow the code to run on several available computers including the Cray Y-MP and personal computers. The capability for modeling attack definitions was increased by incorporating stick-release modes for bombs, multiple aim points, and an increased definition of weapon system error. The code was also modified to handle steel components, and new routines were added for calculation of probability of kill statistics for POL storage and similar facilities.

Target Description

A protective structure or facility is modeled as a combination of critical and noncritical components which define the structures, personnel, and equipment that are relevant to the vulnerability analysis. Each portion of the target can be modeled in great detail with a large number of small components or be crudely modeled with a small number of large components. Components not critical to the function of the facility are often crudely modeled. They seldom require a detailed damage assessment but may add protection through shielding to the critical components of the facility.

Once the target has been divided into the components which will represent it in the vulnerability analysis, the components are given a unique component number and are defined in a manner which can be interpreted by the code. The surfaces of each component are defined by a series of adjacent triangles that have vertices which are given sequence numbers and x, y, and z coordinates. The code stores the triangular surface descriptions and the maximum and minimum x, y, and z values for each component and later uses this information in ray intersection calculations. The target description process can be simplified by using component-generating programs that prepare the required triangular description data for simple shapes.

Vulnerability Criteria

The entire target vulnerability is assessed from the vulnerability of the components that are critical to the target's function. Each component of the target is assigned a material number which identifies its material properties. This material classification determines which damage and vulnerability calculations are used for that component. COBRA uses existing routines for calculation of the damage and vulnerability of components which are classified as parts of people, concrete, wood, steel, aluminum and foam composite, and different types of soil. Components may be classified as other types of materials, and the material properties, damage criteria, and vulnerability data can be entered directly in the input file. The vulnerability data for the undefined materials can be entered as algebraic equations or by using probability of kill versus fragment velocity curves for several fragment weights. This allows the option of determining component vulnerability from raw test data or from methods already developed for the particular material.

The relationships defining the impact of the individual component vulnerabilities to the vulnerability of the entire target are defined in the system definition using Boolean logic. The critical components that make up the target are grouped according to function into critical subsystems. These subsystems can then be grouped together to form systems which are essential to the operation of the targeted facility. An example of this type of system definition can be shown for a chemical

decontamination shelter. The critical components of the decontamination unit would be grouped together to form a subsystem, while the structural components would form another subsystem. The combination of the two subsystems form the system which represents the major functions necessary for proper operation of the decontamination facility.

Attack Definition

A large number of parameters must be input for the attack definition. The projectile parameters include a fuzing classification of contact, delay, or proximity and a nose shape factor. The blast damage potential of the projectile is specified by explosive weight, in pounds of TNT, that would induce a blast pressure equivalent to that of the projectile. The fragment damage potential is determined by the velocity, number, and weight of the fragments in each angular band defined about the nose of the projectile. The program assumes that the fragments in each weight class are uniformly distributed about the nose band. The weapon system errors are defined for the expected threat and can be separated into the precision error, the target location error, and the mean point of impact error. The target aim point can be specified as either a single aim point or as multiple aim points which more accurately model some attack scenarios. In modeling air to surface threats, the projectile release method can be specified as either single or stick release. Additional parameters required to describe the attack definition include the incoming projectile angle, impact velocity of the projectile, the number of rounds in an attack, and the number of attacks.

Attack Simulation

The attack simulation begins by calculating the path of the projectile to the target. The code uses the aim point coordinates, the error ellipse produced by the system errors, and two pseudorandom numbers to calculate the coordinates of the impact point. The computer-generated pseudorandom numbers are converted from a uniform distribution to a normal distribution so that the burst points will be normally distributed inside the error ellipse. The projectile is delivered to the burst point along an entry path with a specified azimuth and incoming angle. The projectile is traced in short increments from the edge of a user specified target hit-miss box along the entry path until it encounters a component or completely misses the target. The projectile velocity remains the same during movement along the entry path since air drag is ignored.

The fuzing condition of the projectile determines which calculations are conducted when the projectile encounters a component. Projectiles which have contact fuzing are detonated at the point where the projectile encounters the component. Proximity-fuzed projectiles are detonated at the proximity distance from the point where the component is encountered. Delay-fuzed projectiles are assumed to operate only in soil or in air after

encountering soil. The delay fuzed projectile detonates immediately if it encounters components of any other material. The assumption is that most structural elements would slow the projectile, and the delay time would elapse and cause detonation at the structural element. This may not be the case for thin structural components. If the projectile enters soil on the entry path, the fuze delay time begins to accumulate. The projectile path is traced through various layers of soil, and the velocity reductions and time required to traverse the layers are calculated. The projectile path through soil is curved to simulate projectile hooking behavior in soil. If the projectile exits soil into air, it retains its velocity and travels in a straight line. The time required to cross the air gap is accumulated as elapsed fuze time. The projectile is detonated when the elapsed fuze time is greater than the fuze delay time or the projectile velocity falls below 10 ft/sec[1].

The equations used for the projectile path and velocity calculations through soil were derived by assuming that the force acting on the projectile was a drag force F_D, of the Poncelet form, as defined in Terminal Ballistics[2].

$$F_D = m \frac{dV}{dt} = -A (a + bV^2) \tag{1}$$

$$\begin{aligned}
m &= \text{mass of projectile} \\
V &= \text{projectile velocity} \\
t &= \text{time} \\
A &= \text{projectile cross-sectional area} \\
a \text{ and } b &= \text{empirical constants}
\end{aligned}$$

The empirical constants, a and b, were chosen so that the calculated path length would correspond to values given in the National Defense Research Committee (NDRC) report[3]. The constants are a function of the projectile's nose shape, caliber density, and the type of soil the projectile is passing through.

The incremental path length, S, of the projectile penetration through soil can be derived from equation 1.

$$S = S_2 - S_1 = \frac{m}{2Ab} \ln \frac{a + b(V_1)^2}{a + b(V_2)^2} \tag{2}$$

Rearranging the terms in equation 2 gives the velocity, V_2, at the end of the incremental path length.

$$(V_2)^2 = \frac{a + b(V_1)^2 - a\, e^{(2Ab/m(\Delta S))}}{b e^{(2Ab/m(\Delta S))}} \tag{3}$$

$$\Delta S = S_2 - S_1$$

After determination of the incremental velocity, the elapsed time can be obtained by direct integration of equation 1.

$$t_2 - t_1 = \frac{m}{A\sqrt{ab}} \left[\arctan V_1 \sqrt{(b/a)} - \arctan V_2 \sqrt{(b/a)} \right] \qquad (4)$$

The equations used to determine the orientation of the projectile and its deflections during soil penetration are similar to equations in "Terminal Ballistics"[2]. The equation assumes the deflecting force is a component of the total drag force and relates it to a centripetal force on a body in a curved trajectory. The instantaneous deflection is determined in an angular form and then converted to a change in coordinates. The derivation of the equation used for calculation of the instantaneous deflection can be found in Cheek's report[1]. The coordinate changes due to the projectile deflection are given by equations 5 - 7.

$$X_2 = X_1 - \Delta S \cos \alpha \sin \beta \qquad (5)$$

$$Y_2 = Y_1 - \Delta S \sin \alpha \sin \beta \qquad (6)$$

$$Z_2 = Z_1 - \Delta S \cos \beta \qquad (7)$$

X_1, Y_1, Z_1 = previous coordinates
X_2, Y_2, Z_2 = new endpoint coordinates
ΔS = segment length
α = azimuth angle
β = average angle that the trajectory of projectile makes with the z axis (includes instantaneous deflection)

After the location of the burst point has been determined, it is necessary to determine which components are affected by the projectile detonation. This is accomplished by generating rays which project from the burst point in a spherical pattern. The program calculates the intersection of the ray with each component it encounters. It stores the components on each ray in increasing order with respect to the distance of the component to the burst point. The program can then subject the nearest components to the blast and fragmentation associated with the projectile detonation. If the components nearest the detonation are defeated, the blast and fragmentation effects are then applied to the next component on the ray.

The best results in the damage calculations will be obtained if at least one ray passes through every critical component of the target. There are two ray densities, large grid and small grid, which can be specified in modeling the interaction of the component with the blast and fragmentation

of the projectile detonation. The large grid specification will result in fewer rays being passed through the component and is usually adequate for large components with a simple shape. The small grid specification is used for small components or large components with complex shapes and attempts to force at least one ray through the smaller areas.

The large and small ray grids are developed by dividing a unit sphere into a number of sections determined from a user defined value for the grid size as shown in equation 8.

$$D = 2 N \{[(N - 2) / 4] + 1\} + 2 \qquad (8)$$

$D =$ number of divisions for unit sphere
$N =$ user specified value for large or small grid size

The specified value of N corresponds to dividing a circle (360 degrees) into N equal parts. The elevation increment of the large and small grid divisions consists of E equal parts of $\sin \theta$ while θ varies from -90 to +90 degrees. The value for E is calculated using equation 9.

$$E = 2 \{[(N - 2) / 4] + 1\} \qquad (9)$$

$E =$ number of elevation increments on unit sphere

Additional rays are produced along with the grid rays which pass through the centroid of each component of the target. The centroid rays assure that at least one ray will pass through all the critical components of the target.

A ray selection process is used to identify the rays which pass through critical components having the same grid specification of either large grid or small grid. If the ray does not insect with a component having the same grid specification, it is deleted. Another criteria for the selection of a ray is that the triangular surface element which is intersected must be larger than the grid element of the ray. This eliminates the rays which completely miss the target or only impact with small parts of a component. If the ray generated through the centroid of a component or selected from the large grid rays are in approximately the same location as a ray selected from the small grid rays, then only the small grid ray is used in the calculations. The surfaces intersected by the other rays are transferred to the small grid ray.

Vulnerability and Damage from Blast

Blast induced damage is simulated using blast propagation models which calculate the effect on various structural elements. The vulnerability of the component is then calculated with regard to the possible blast and fragmentation from previous rounds.

The ground shock force on a face or object of solid angle ω can be calculated using a pressure-distance empirical relationship written as shown in equation 10.

$$p = \frac{kW}{r^3} \tag{10}$$

k = soil constant
W = charge weight
r = range to detonation

Since the area of solid angle ω is $\omega\, r^2$, the force on the object can be calculated using equation 11.

$$F = \frac{k\, W\, \omega}{r} \tag{11}$$

When blast propagates through soils with varying k values, the thickness of the second soil layer is adjusted to be an equivalent thickness in the first soil layer using a ratio of the soil constant for the first type of soil and the second type of soil. This method ignores reflections between soil layers and air gaps in the soil of less than 1 inch. It assumes no blast propagation across air gaps larger than 1 inch. The pressure equation used in the ground shock model is shown in equation 12.

$$p = (1 + \cos\theta)\,\frac{k}{2\lambda^3} \tag{12}$$

θ = angle of obliquity
λ = scaled distance, $ft/lb^{1/3}$

An air shock model based on standard atmospheric pressure and temperature is also included. Since the shock model for this code was developed for underground bursts, the blast pressure from air is only transmitted to the structural elements of the target and not to the soil layers. The force transmitted to the structural elements is applied not only to the first element on the ray but also the elements which are immediately touching the first element. Air gaps of less than 1 inch and soil gaps of less than 5 inches between elements are ignored. The pulse length is set at 0.001 second. The pressure equation used in the air shock model is shown in equation 13.

$$p = \frac{1 + \cos\theta}{2} * \left[\,\frac{4120}{\lambda^3} - \frac{105}{\lambda^2} + \frac{39.5}{\lambda}\,\right] \tag{13}$$

λ in inches / $lb^{1/3}$

The blast pressure applied to each structural element from the different shock models is added to a force vector for that element. The accumulated forces are then used to calculate blast damage. The structural components of the target may consist of several types of material with different failure modes; therefore, several methods for calculating the blast damage and vulnerability of a component have been included. The most commonly used methods are for components consisting of reinforced concrete or materials which deform elastically; however, if the component consists of a material which cannot be evaluated with these methods the vulnerability equation can be entered directly in the input file.

Materials which deform elastically until failure are treated as harmonic oscillators responding to a forcing function. For blast, the forcing function approximates a triangular pulse. Based on the pulse duration and peak pressure, the maximum stress applied to the element can be calculated. The blast calculation determines the ratio of the maximum stress due to shock to that which is caused by a static force of the same magnitude. After the ratio is applied to the static value, a comparison is made with the maximum possible loading on the structural element. A probability of kill is produced which states that the element has either failed, not failed, or possibly failed. The probability of kill is combined with the previous values to produce a cumulative probability of kill.

To utilize this method, the outward facing area, length, maximum bending moment, and period of vibration for each component must be entered in the input file. The equations for calculation of the maximum bending moment (14) and period of vibration (15) are shown below.

$$M_c = \frac{I\,\sigma}{c} \tag{14}$$

M_c = maximum or critical bending moment
I = moment of inertia of the cross section
σ = stress at failure
c = maximum distance from the neutral axis

$$T = \frac{2L^2}{\pi} \sqrt{(\gamma A\ /\ EIg)} \tag{15}$$

T = period of vibration
L = length
γ = specific weight
A = cross-sectional area
E = Youngs modulus
g = acceleration of gravity

The applied bending moment is approximated from the triangular force applied to the element as shown in equation 16.

$$M = \frac{F\,L}{8}\qquad(16)$$

M — applied moment
L — length of the element

The probability of kill for the component is determined based on the comparison of the applied bending moment to the maximum bending moment. The relationship between these comparisons and the probability of kill is shown below.

$$P_k = 0.0 \quad \text{if} \quad M < 0.8\ M_c$$

$$= 0.5 \quad \text{if} \quad 0.8\ M_c \leq M < 1.2\ M_c$$

$$= 1.0 \quad \text{if} \quad M \geq 1.2\ M_c$$

Reinforced concrete is treated with tabulated data obtained from the National Defense Research Committee report[3]. The tabulated data was compiled for reinforced-concrete walls buried in soils with a soil constant of 5000. It also assumes that the ground shock is propagated from a charge buried opposite the center of the wall. Adjustments are made to the calculations to allow for circumstances which are not directly covered by the tabulated data. The thickness and the area of each component are required for this method.

The total force applied to the element and the area of the element are used to calculate the average pressure acting on the component. The average pressure is then increased by 30 percent to obtain a peak pressure consistent with the values used in the NDRC report[3]. The peak pressure is converted to a scaled distance using an equation of the following form.

$$\lambda = (k\ /\ P)^{1/3}\qquad(17)$$

λ — scaled distance
k — soil constant
P — peak pressure

The scaled distance is related in the tabulated data to a degree of damage and probability of kill. The probability of kill more closely represents a degradation of function due to spalling, cracking, and crunching than a probability of failure.

The cumulative damage to structural components from the effects of blast is represented by the cumulative probability of kill for the

component. The probability of kill calculated for one detonation is combined with the cumulative probability using following equation.

$$P_c = P_1 + P_c - P_1 P_c \qquad (18)$$

P_c = cumulative probability of kill
P_1 = probability from one detonation

Fragment Damage and Vulnerability

The potential for damage from fragmentation to the components of the target is determined by assuming that fragments travel along the rays from the burst point to the components on the ray. The nose band of fragments which the ray passes through determines the fragment weight and velocity pairs that will be used in the damage calculations for the components on that ray. The number of fragments which travel along the ray is calculated using equation 19.

$$N = N_B \ (A_G \ / \ A_B) \qquad (19)$$

N_B = number of fragments in weight group on band
A_B = area of unit sphere covered by the band
A_G = area of unit sphere covered by grid element

The fragment data input to the program is typically determined from static tests; therefore, the program adjusts the input fragment velocity and direction to reflect the projectile velocity and direction. If the distance that the fragments must travel along the ray before intersection with the first component is greater than 2 feet, the velocity of the fragment is adjusted for the effects of air drag. Air drag is ignored if the fragment travels less than 2 feet before intersection with a component. The adjustment to the fragment velocity is calculated using equation 20.

$$V_I = V_0 \ e^{-(0.002363 \ D/W^{1/3})} \qquad (20)$$

V_I = impact velocity
V_0 = original velocity
D = distance from burst point to first component, in.
W = weight of fragment, grains

Once the fragment travelling on the ray intersects with a component it is necessary to calculate the depth of penetration and determine if the fragment perforated the component. Fragments which perforate a component continue to travel along the ray with a reduced velocity until intersection with the next component. Fragment penetration equations are included in

the program for soils, wood, concrete, steel, people, and an aluminum and foam composite.

The equation for fragment penetration into sand or soil is from the Poncelet equation and has the form shown below.

$$D_p = \frac{m}{2Ab} \ln [1 + (b V^2) / a] \qquad (21)$$

D_p = penetration distance
m = fragment mass
A = fragment presented area
V = fragment impact velocity
a and b = empirical constants

The constant terms in this equation were adjusted to match the curves for fragment penetration through sand in an Air Force Weapons Laboratory (AFWL) report[4].

The concrete penetration equation was taken from a paper by A. K. Kar[5] and is shown in equation 22. A ricochet model has also been added to more correctly predict the interaction of fragments with concrete components.

$$P = \frac{0.10237 \times 10^{-6}}{(f_c)^{1/2}} N_2 (E / E_m)^{5/4} \frac{W}{D} (V / d)^{9/5} + 1 \qquad (22)$$

P = penetration in calibers
f_c = compressive strength of concrete, psi
E = elastic modulus of the fragment
E_m = elastic modulus of mild steel
d = diameter of circle with area equal to impact face, in.
D = outside diameter of nonsolid nose shapes, in.
N_2 = nose shape factor, 0.72 for fragments
V = projectile impact velocity, ft/sec
W = projectile weight, grains

Fragment penetration into wood is calculated using the empirical relationship shown below.

$$E = V^c W^d (\sec \phi)^e \eta^f \qquad (23)$$

E = penetration, in.
V = fragment impact velocity, ft/sec
W = fragment weight, grains
ϕ = obliquity angle measured from the surface normal
η = Young's modulus for the impacted wood
c, d, e, f = empirical constants

When the penetration of the fragment is larger than the component thickness an exit velocity for the fragment is calculated. For all components other than steel or people the exit velocity is calculated using the following equation.

$$V_X = V_E (1 - T / E)^{1.8} \tag{24}$$

The exit velocity for fragments passing through people is calculated using equation 25.

$$V_X = V_E - A T W^B (V_E)^C \tag{25}$$

$$
\begin{aligned}
V_X &= \text{exit velocity, ft/sec} \\
V_E &= \text{entrance velocity, ft/sec} \\
W &= \text{fragment weight, grains} \\
T &= \text{line-of-ray thickness, in.} \\
A, B, C &= \text{empirical constants}
\end{aligned}
$$

Fragment damage is evaluated in this code using one of three different methods. The first method is the volume method where the probability of failure of a component corresponds to the amount of cumulative volume that has been removed from the component. The cumulative damage is modeled by adjusting the density of the component according to the volume removed. The second method is the use of algebraic equations to determine a probability of kill based on the fragment weight and velocity. The third method is the use of fragment velocity versus probability of kill curves which are entered for the component type at each fragment weight.

System Probability of Kill

After all the damage calculations have been completed, the program compiles the probability of kill statistics for the subsystems and systems in the target vulnerability criteria. When components and subsystems are combined with an "and" Boolean operator, the subsystem definition implies that all functions represented must be killed for the subsystem probability of kill to be 1.0. When the probabilities are combined using the "and" operator the following equation is used to calculate the probability of kill.

$$P_k(C) = P_k(A) \, P_k(B) \tag{26}$$

$$
\begin{aligned}
P_k(C) &= \text{combined probability of kill} \\
P_k(A) &= \text{probability of kill for function A} \\
P_k(B) &= \text{probability of kill for function B}
\end{aligned}
$$

Components and subsystems can also be combined with an "or" Boolean operator which implies that if any one of the functions represented are

killed, the subsystem probability of kill will be 1.0. Equation 27 is used to calculate the combined probability if the "or" operator is used.

$$P_k(C) = 1.0 - [1.0 - P_k(A)] [(1.0 - P_k(B)] \qquad (27)$$

This entire process is repeated until the user specified number of projectiles have been examined or the defined system vulnerability reaches a cumulative probability of kill of 1.0. The cumulative probability of kill is determined for each subsystem and system in the target vulnerability criteria using equation 27.

Conclusions

Performing a vulnerability analysis for a facility using the probabilistic methods in COBRA allows the user to define the critical aspects of the facility and determine the chances of survival for particular threats without a large expenditure for field testing. The detailed damage assessment for the vulnerable areas of the facility allow the designer to determine which types of protection schemes may be the most appropriate, and after another vulnerability analysis it is possible to determine if certain areas of the facility still suffer unacceptable levels of damage. Since the vulnerability analysis is cost effective, it is possible to evaluate several protection schemes and determine which one provides the highest level of protection. Cost comparisons versus level of protection can then be obtained.

This type of analysis was recently performed for an Army POL Storage Site. The vulnerable components of the site were analyzed for two attack definitions using the probabilistic vulnerability methods in the COBRA code. The probability of kill for the storage site was compared and agreed favorably with values calculated by others. Several protection schemes were then developed, and the vulnerability analysis was conducted for each scheme. The information obtained from the vulnerability analysis allowed comparison of the level of protection provided by each protection scheme and the cost of implementing the protection method. The probabilistic vulnerability analysis has provided an extremely cost effective method which can be used to analyze many protection schemes and determine which method should be used to provide the most cost-effective level of protection.

Acknowledgments

References

1. Cheek, James B., Jr., "Development of a Program for Computer-Based Resistance Analysis (COBRA)," Technical Report SL-80-5, U.S. Army Engineer Waterways Experiment Station, Vicksburg, Mississippi, July 1980.

2. Backman, M. E., "Terminal Ballistics," NWC TP 5780, Naval Weapons Center, China Lake, California, February 1976

3. National Defense Research Committee, "Effects of Impact and Explosion," Vol I, Washington, D.C., 1946.

4. Crawford, R. E., et. al., "Protection from Nonnuclear Weapons," TR-70-20-120, Air Force Weapons Laboratory, Kirtland Air Force Base, Albuquerque, New Mexico, February 1971.

5. Kar, A. K., "Local Effects of Tornado-Generated Missiles," Journal of The Structural Division, American Society of Civil Engineers, Vol 104, No. ST-5, May 1978, pp 809-816.

Optically Controlled Dielectric Resonator Oscillator for Millimeter Wave Applications (U)

*Thomas P. Higgins, Mr. and Arthur Paolella, Mr.
U. S. Army, LABCOM,
Electronics Technology and Devices Laboratory,
Ft. Monmouth, N. J. 07703

Peter R. Herczfeld, Dr.
Drexel University,
Center for Microwave- Lightwave Engineering,
ECE Dept.,
Phila., Pa. 19104

INTRODUCTION

New Monolithic Microwave Integrated Circuit (MMIC) technology has the potential to greatly reduce cost and size as well as improve performance of microwave/millimeter wave systems. Other advances have made millimeter wave circuits viable candidates for ground and satellite systems. Some of these systems will utilize active phased array antenna concepts, now being investigated at microwave frequencies, to steer and shape the antenna's radiation pattern. These techniques enable target tracking, in the case of radar, and beam pointing to select specific ground based receivers for communications from satellites. Active phased array antennas are made of many closely spaced (one-half the operating wavelength) individually controlled Transmit/Receive (T/R) modules, each with its own radiating element. The distribution of control and synchronization signals (phase shifting, gain control, switching, locking, ect.) to microwave and millimeter wave modules, for communications and radar systems, is a formidable task requiring large size metal waveguide or coaxial transmission line distribution networks. Many of these systems will require remote location of antennas, transmitters, and receivers. Remote location of microwave circuits and systems is limited due to losses of conventional transmission lines and their susceptibility to EMI. Interconnects between chips also have high loss, and radiation problems at microwave frequencies. To further reduce size, in many cases it may be advantageous to replace RF/Microwave interconnects with electro-optic techniques. Among the advantages of using optical fiber cables are lightweight, small size, broad bandwidth, immunity from EMI and very high isolation between adjacent cables.

In sophisticated systems requiring low noise, highly stable oscillators, Gunn oscillators are used with dielectric resonators. Dielectric resonators have a very high Q, which locks

HIGGINS, PAOLELLA, HERCZFELD

the frequency, and with the proper choice of temperature coefficient, can result in nearly constant frequency with changing temperature. The traditional method of tuning a DRO is mechanically with a plunger inside the resonator cavity. The movement of the plunger closer to the dielectric resonator changes the boundary conditions imposed on the fields inside the resonator cavity, which shifts the oscillation frequency upward. Electrically, there exits two methods of electrically tuning this type of oscillator. The first is by shifting the bias voltage (voltage pushing), which is very limited and not always practical because of the variation in output power; and the second method is by the incorporation of a varactor diode into the circuit. Varactors reduce the circuit Q, as well as complicate the circuit design and construction. Fiber optics can be used to provide remote tuning while reducing the effects of EMI.
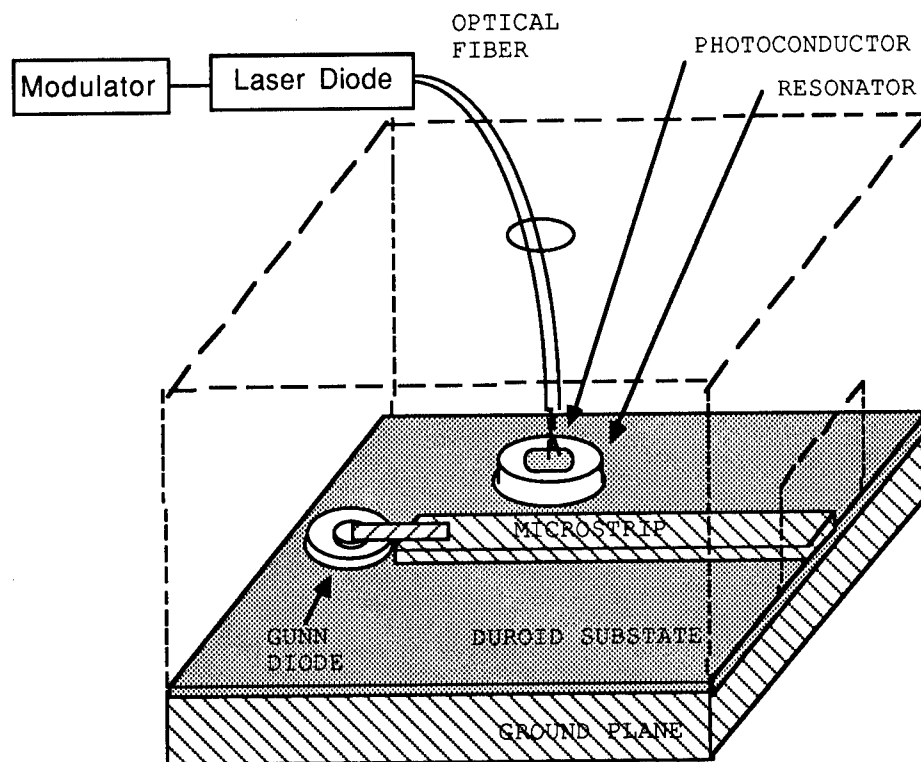
Fig. 1—Optically controlled DRO concept.

For this investigation, a Ka-band Dielectric Resonator Oscillator (DRO) using a Gunn device, was designed, and the successful replacement of the mechanical tuner with an optical tuner was achieved. The oscillator was designed such that a photoconductor (semi-insulating GaAs) could be placed on the resonator, and an optical fiber inserted into the oscillator cavity above the photoconductor. An optical signal emitting from the fiber creates electron-hole pairs in the GaAs, which increases the conductivity, thus perturbing the fields

HIGGINS, PAOLELLA, HERCZFELD

as if the mechanical tuning plunger was moving closer, thereby raising the resonant frequency. The frequency of oscillation and FM modulation was controlled via a fiber optic link to demonstrate the viability of using optics as a distributive network.
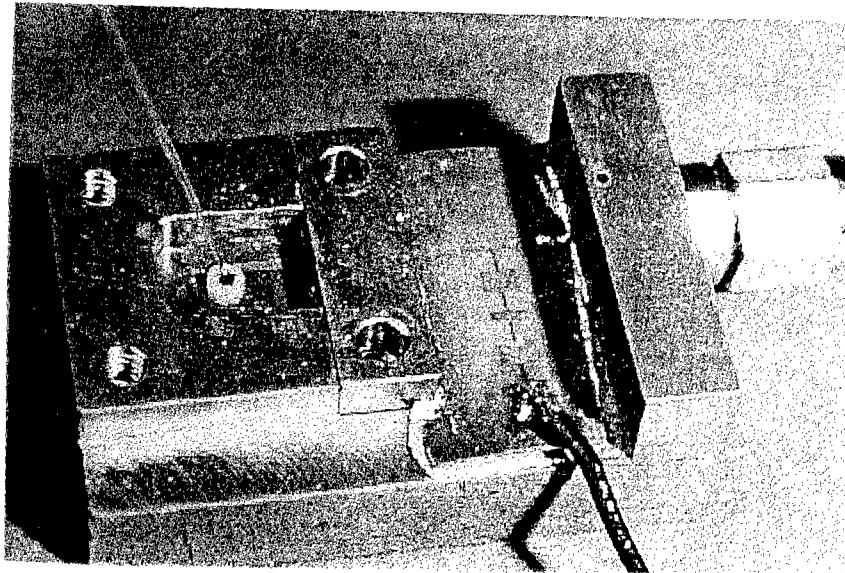
Fig. 2—DRO with GaAs sample and optical fiber

Dielectric Resonator Design and Evaluation

The construction of the DRO is, as shown in Fig. 1, fabricated on a 0.010 inch thick duroid microstrip bonded to a grounded heatsink. The packaged Gunn diode was screwed into the heatsink inside the enclosure with the top contact, ribbon bonded to the end of the 50 ohm microstrip line. A Molecular Beam Epitaxy (MBE) Gunn device was used, which had a threshold voltage of 1.83 volts, threshold current of 1.08 amps, and produced 210 milliwatts of power at Ka-band, in a waveguide test cavity at 5.1% efficiency. This device was designed with its active region impurity profile graded to improve its thermal characteristics and increase its output power[1] . The d.c. bias to the device was provided by a three-section distributed low pass filter on microstrip (not shown). The filter was designed using Touchstone and has greater than 25 dB attenuation from 25 to 40 GHz. The RF output was to a K Connector. A stub for mechanical tuning of the operating frequency was incorporated into the top cavity wall and can be removed for easy insertion of the optical fiber. The cylindrical dielectric resonator, made by Murata-Erie, consisted of a $Ba(TiTa)_3$-$Ba(ZrZnTa)_3$ compound with a dielectric constant of 29, and an unloaded Q, of 3200 at 28 GHz. The resonator had a diameter of 2.19 mm and a height of 0.97 mm, which

HIGGINS, PAOLELLA, HERCZFELD

allows for operating frequency from 27.5 to 30 GHz. The height to diameter ratio is chosen for good Q, and good separation from other modes. The cavity dimensions were chosen, so as to be large enough to allow for experimentation in the placement of the resonator for achieve maximum power output, and maintain the minimum one resonator diameter distance from resonator and cavity wall. The resonator was placed very close to the 50 ohm microstrip line, approximately one quarter wavelength from the diode; in this way the resonator also acts as an impedance transformer, matching the diode to the circuit [2]. When the position for maximum RF power generation was found, the resonator was epoxyed into place.
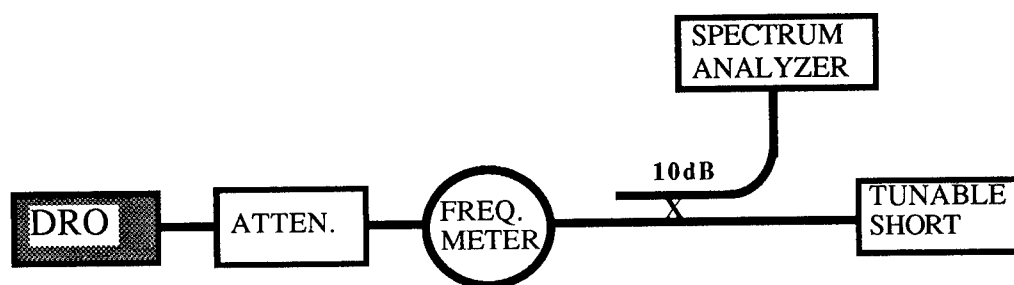
Fig. 3—Experimental setup to measure Q of DRO.

Once constructed, the oscillator was evaluated for power and frequency. The DRO produced 35 mw of rf power and was mechanically tunable from 28.7 to 29.8 GHz which was within 0.35% of the value predicted by Murata-Erie's design software called "Resomics". Next, loaded Q factor measurements were taken using the method described by Warner and Hobson [3], and shown in Fig. 3, which revealed a Q of about 730. This can be adjusted by changing the distance between the resonator and the microstrip line; the further away the resonator is, Q is increased, however output power is decreased. The measurement was made with a movable discontinuity provided by a tunable short in the end of the test line with an in-line attenuator set to provide the required return loss. The line was calibrated so that VSWR of 1.5 could be accurately set. The peak-to-peak frequency variation was measured on the spectrum analyzer, then the loaded Q can be calculated by:

$$Q_L = \frac{f_o Q_o \left(s - \frac{1}{s} \right)}{2 Q_o \Delta f + f_o \left(s - \frac{1}{s} \right)}$$

where VSWR is represented by s and operating frequency of the oscillator is fo.

The DRO was made optically sensitive by first removing the mechanical tuner, inserting the fiber in its place, and placing a photoconductor on top of the dielectric resonator[4]. The photosensitive material used, were samples of high resistivity GaAs (10^8 ohm-cm dark

HIGGINS, PAOLELLA, HERCZFELD

resistivity), and ground to a thickness of three mils, so as to affect the resonator Q minimally. This optical port allows the photoconductor to be illuminated by various light sources while the cavity cover still contains the fields. When light illuminates the photoconductor with energy greater than the band gap of GaAs, electron-hole pairs are generated, causing a change in the conductivity of the photoconductor, thus changing the boundary conditions imposed on the fields, which shifts the oscillation frequency. Optical tuning and modulation of the oscillator was obtained by illuminating the sensitized dielectric resonator through the hole in the top of the cavity and the effect recorded on a spectrum analyzer as shown in Fig. 4.

Experimental Procedure and Results

Frequency tuning experiments

It was found that when the GaAs samples were placed on the resonator, the output power of the oscillator decreased, while shifting the operating frequency downward slightly (from 28.7 to 28.69 GHz). Light incident on the resonator would shift the center frequency of the oscillator upward. The reaction of the oscillator to three different light sources was examined. A strong white light source was used to determine the sensitivity of the samples.
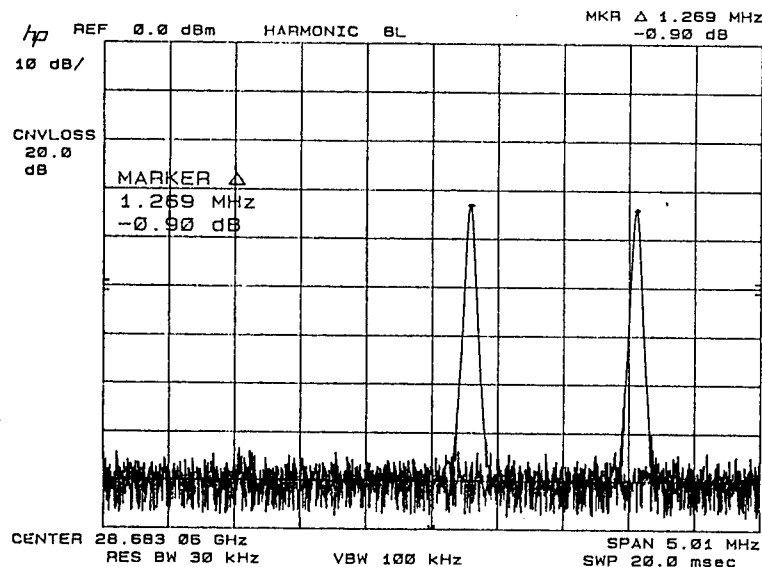
Fig. 4—Shift in resonant frequency with white light. The left trace is with no illumination, the right trace with illumination.

HIGGINS, PAOLELLA, HERCZFELD

Samples ranged in size from $0.1mm^2$ to $0.6\,mm^2$. The largest size samples produced the most frequency change when illuminated, but reduced the output power of the oscillator. Smaller samples produced a shift of 0.1 MHz with 23 mw of rf power; medium sizes produced 0.6 MHz of change with 12 mw of rf power; larger samples produced as much as 5 MHz change with 4 mw of rf power. The oscillator's sensitivity to light was highly subject to the placement of the photoconductor on the resonator and the size of the sample. A He-Ne gas laser at 632 nm wavelength with an output of 1 mw was used, and shifted the operating frequency 1 MHz with 1 dB decrease in output RF power. An RCA 86006E laser diode was used which produces a maximum 4 mw of peak optical power into an optical fiber at a wavelength of 820 nm was then used which demonstrated a 5 MHz tuning range. Testing reveals a linear relationship as shown in Fig 5, between incident optical power and shift of oscillating frequency, with the maximum change of 5 MHz from 4 mw of optical power. The results approach what has been achieved using electrical tuning methods.
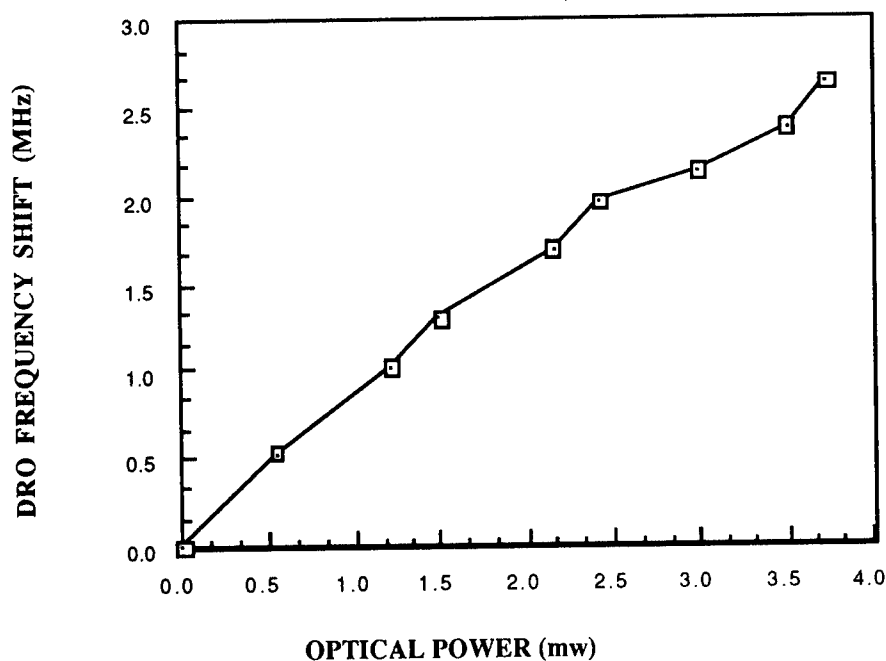
Fig. 5—Shift in the DRO resonant frequency as a function of IR Laser output power.

FM modulation experiments

The design of the optically controlled DRO allows FM modulation of the oscillator by AM modulating the input optical signal. To achieve this, a mechanical chopper was used to modulate a He-Ne laser with a 50% duty cycle pulse, which was directed onto the
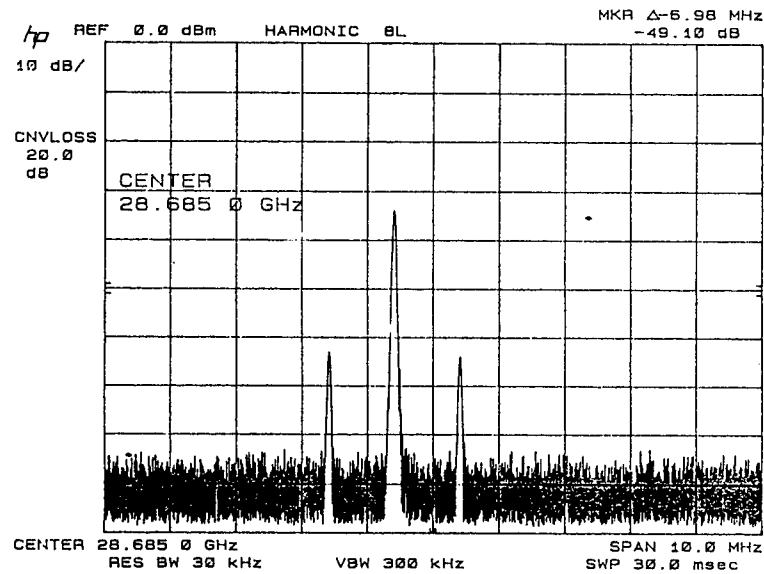
Fig. 6—Narrow-band FM spectrum with IR Laser illumination.
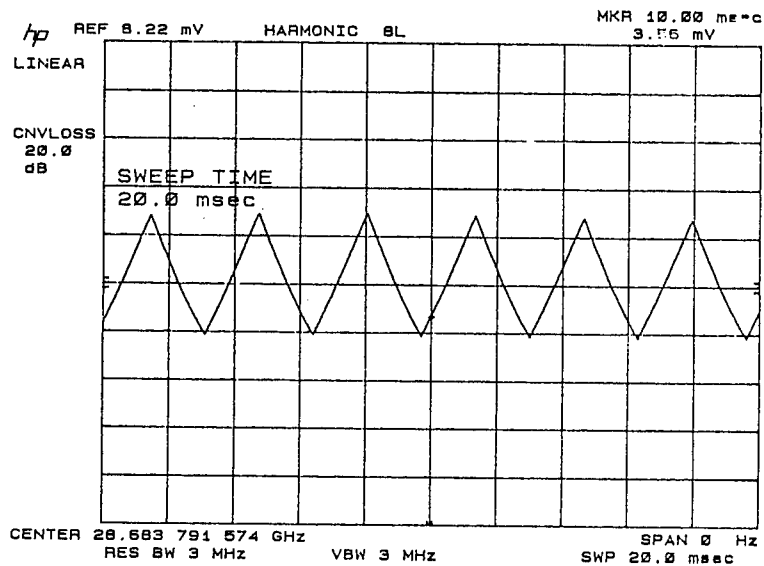
Fig. 7—Demodulated RF reveals original signal.

HIGGINS, PAOLELLA, HERCZFELD

photoconductor. The maximum modulation rate achieved was limited by the chopper which was 4 KHz. The DRO was then modulated with a laser diode, at a maximum frequency of 7 MHz, limited by the laser. The frequency spectrum with the modulation sidebands can be seen in Fig. 6 at 1 MHZ. In addition the laser was modulated with a triangle wave; the RF could then be demodulated to reveal the original signal to show linearity as shown in Fig. 7.

DISCUSSION

We have presented a technique to control the operating frequency of a Millimeter Wave Dielectric Resonator Oscillator by optical means. A MMW DRO was design and fabricated in a MIC configuration using a MBE Gunn device as its active element. The oscillator was designed such that a photoconductor could be placed on the dielectric resonator and a fiber could be inserted into the oscillator cavity above the photoconductor. Light emitting from the fiber changed the conductivity of the photoconductor thus, producing a change in operating frequency of the DRO . A maximum change in the oscillating frequency of the DRO of 5 MHz was obtained with 4 mw of optical power from a diode laser with modulation rates up to 7 MHz. The oscillator design implemented in this manner enables frequency modulation information to be sent to the oscillator via the fiber optic link.

Applications of optically controlled DROs include the distribution and synchronization of the MMW signals via fiber optic links to the large number of closely spaced individual T/R modules utilized in active phased array antenna concepts, now being investigated at microwave frequencies, to steer and shape the antenna's radiation pattern.

REFERENCES

1. Paolella, A., Ross, R. L., and Ondria, J., " Advanced mm-Wave Sources by MBE", *Microwave Journal,* Vol. 29, No. 4, April, 1986.

2. Makino, T., and Hashima, A., "A Highly Stabilized MIC Gunn Oscillator Using a Dielectric Resonator", *IEEE Transactions on Microwave Theory and Techniques,* Vol. MTT-27, No. 7, July 1979.

3. Warner, F. L., and Hobson, G. S., "Loaded Q Factor Measurements on Gunn Oscillators", *Microwave Journal,* Vol. 16, No. 2, Feb., 1973.

4. Herczfeld, P. R., Daryoush, A. S., Rosen, A., Stabile, P., Contarino, V. M., "Optically Controlled Microwave Devices and Circuits", *RCA Review,* Vol. 46, December 1985.

Does OPTEMPO Increase Unit Readiness?
An Objective Answer (U)

*Dr. Jack H. Hiller
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Howard H. McFann
ARI Presidio of Monterey Field Unit
Monterey, CA 93940-5011

BG Lawrence G. Lehowicz
Office of the Deputy Chief of Staff,
Operations and Plans - Training, Pentagon
Washington, DC 20310-0200

The purpose of this research was to design and test an operationally practical methodology for determining if level of OPTEMPO relates to objectively measured unit performance capability. The hypothesized relationship between military resources expended in field training by units and objectively measured gains achieved in their "training readiness" for combat is a venerable issue for the Congress, the Department of Defense and its military departments. The issue has become critical while a large Federal budget deficit continues and the threat of a major war appears to be diminishing. A direct linkage between training time/activity and reasonably objective ratings of proficiency has in fact been demonstrated to some extent for:

a) Navy aircraft flying hours with performance measured by bombing accuracy and landing success rate (DePoy[4], and Horowitz, et. al.[6]);

b) Air Force pilot bombing accuracy (Cedel and Fuchs[3]).

However, the Army has traditionally assessed the training readiness or proficiency of its ground forces by having unit commanders subjectively rate their unit's capabilities and estimate the additional training

necessary to be prepared for combat. The use of subjective estimation to establish the proficiency of ground forces derives from three major measurement difficulties:

a) The Multitude of Missions & Tasks. There is no single, or even a few, unit tasks that can be selected as a focus for objective measurement; instead there are tens of complicated missions (e.g., Deliberate Attack, day or night, Defend in Sector, and Movement to Contact) involving hundreds of tasks performed at numerous echelons from infantry fire teams and squads, through platoon, company, battalion, brigade, division, and possibly corps and theater army. There is so much to measure that experienced military judgement has provided the only practical basis for estimating unit capability and training needs.

b) The Uniqueness of Each Unit and its Post. Whatever unit performance might be selected for measurement, for the measures to apply to routine operational training (instead of a grand experiment that might be arranged, and thereby fail to represent normal unit training and pro-ficiency), they would have to be implemented at each unit's home-station training grounds. Since the terrain varies significantly from one Army post to another and the units often have distinctly different missions, and since the use of opposing forces to create realism is highly variable at each post, and no posts have the instrumentation required to support an objective, standardized measurement system -- standardized unit per-formance data could not be collected without a major investment, that would likely fail to overcome the inherent differences across posts.

c) Difficulty of Measuring Leadership and Cohesion. The intangible qualities of leadership and unit cohesion that are critical to combat are not directly, easily measurable by unit outsiders, but must be assessed within the units on a subjective basis.

Despite these measurement difficulties, the Army was challenged to demonstrate with objective measures of unit performance, instead of expert judgement as traditionally used, that the level of expenditures for operating heavy equipment during training (i.e., Ground Operating Tempo, or OPTEMPO) relates to combat effectiveness.

Solution Requirements and Constraints

The Assistant Secretary of Defense for Force Management and Personnel requested the Army to provide a plan for justifying the value of OPTEMPO within a few months and the results within a year's time. Statistical

172

requirements for a representative sample of heavy units and repeated performance measures of each sampled unit, to provide a reliable assessment of capability, obviously conflicted with the requirement to produce results quickly and a practical need to avoid interfering with unit training schedules to perform an experiment.

The National Training Center (NTC), Ft. Irwin, CA provides the most advanced, realistic simulated combat training for heavy forces in the free world. A critical feature of the NTC is its instrumentation which tracks the position of friendly and opposing force (OPFOR) weapon systems and records time-tagged vehicle hits and kills based on the use of the Multiple Integrated Laser Engagement System (MILES) which substitutes laser beams for munitions. The NTC provides training on a core set of missions in essentially the same terrain with an OPFOR that is reliably effective. The Army Research Institute maintains a comprehensive Combat Training Centers research archive for all of the simulated NTC battles at its Presidio of Monterey Field Unit. The availability of this extensive data base for heavy unit training conducted at the NTC provided the basis for a solution (see Hiller[5], for a discussion of the research opportunity created by the NTC).

METHOD

Criterion Variable

The performance effectiveness of heavy units (specifically the Battalion task forces composed of mechanized infantry and armor units) was conceptualized as having three principal dimensions:

a) Attrition of enemy forces (OPFOR);

b) Survival of friendly forces (BLUFOR);

c) Control of terrain.

The data were available in the CTC research archive for determining OPFOR attrition and friendly force survival (for instrumented weapons systems, principally tanks and armored personnel carriers), but objective determination of terrain control at the declared end of a battle was problematical. Therefore, the criterion performance variable was constructed as a traditional casualty exchange ratio (CXR):

$$CXR = \frac{\text{Total OPFOR vehicles killed/Number of OPFOR vehicles}}{\text{Total BLUFOR vehicles killed/Number of BLUFOR vehicles}}$$

Data for over ten thousand MILES rounds fired in 58 defensive missions and 42 offensive missions were used to calculate CXR values.

Since the data analysis requires pairing the unit's training mileage with its performance, as computed by the CXR, the unit performance data from the NTC had to be reorganized. Although the tank mileage data were collected from each tank battalion as a single unit at home station, the tank battalion is task organized into two battalion task forces while training at the NTC. Therefore, the criterion performance measures were formed by first calculating the CXR for each of the two battalion task forces at the NTC on each mission. The CXRs for a given task force on defense were averaged (for the 3-5 missions fought), as were the CXRs for the sister task force; and the two Mean CXRs for defense were then averaged. A Mean CXR was likewise calculated for the offensive missions.

Predictor Variables

Three predictor variables were hypothesized to be important: first, tank mileage; second, the similarity of a unit's home-station training areas to the NTC; and third, personnel stability. It was found that all units created and maintained a high level of fill in preparation for the NTC (mean = 86% on a quarterly measure), so that there was no effective variability in personnel stability (s.d. = 4%); hence stability was dropped as a possible performance predictor.

Tank Mileage. Monthly reports for the mileage of each tank in a unit were available for the six months preceding a unit's training at the NTC (typically, the period of maximum preparation) from the Army Materiel Command's Materiel Readiness Support Activity data base that had been built to support tank maintenance. Using these data, the Mean tank mileage per month, per battalion was calculated.

Similarity to NTC. Three individuals, who were highly familiar with the NTC and the home-station training areas of the units sent to train at the NTC, were independently asked to rate similarity on a five point rating scale. As preparation for a summary rating of similarity, ten questions were asked about the nature of home-station terrain, its maneuverability, and availability for unit training. The ratings of the three individuals were found to be in complete agreement for the summary rating.

174

<u>Unit Sample.</u> At the point that research was initiated, there were data available from three years of training at the NTC by 16 brigades from five posts. In addition, the NTC OPFOR were treated as a home station unit for certain analyses; the OPFOR provide data representing an extreme for tank training mileage, given their intensive year long schedule of participation at the NTC.

<u>Hypothetical Unit Performance Model.</u>

Based on the great body of practical experience and the research literature on the relationship between number of practice trials (see for example Bryan and Harter[1]) or time on task (Carroll[2]), we may expect a slowly rising performance curve as practice, indexed by tank mileage, begins and increases, then at some point a fairly steep rise, and finally an asymptotic plateauing, as shown in Figure 1. Hypothetical Relationship of Unit Performance to OPTEMPO. Based on the expected curvilinear relationship demonstrated in Fig. 1., logarithmic transformations were applied to the tank mileage and CXR values prior to performing linear regression and correlational analyses.

<div align="center">RESULTS</div>

<u>OPTEMPO Correlations.</u>

The Pearson product moment correlation between Defensive CXR and tank mileage was .69 with the OPFOR included ($N = 17$, $p < .002$, see Figure 2. Regression of Log Defensive CXR on Log Tank Mileage); and without the OPFOR, the correlation was .64 ($N = 16$, $p < .01$). However, although on Offense the Mean OPFOR CXR was significantly higher than the Mean BLUFOR CXR (1.2 vs .86, $t = 3.2$, $N = 15$, $p < .01$), there were no significant correlations between the CXR and tank mileage because of the extreme variability in the BLUFOR performance; that is, shear amount of BLUFOR training was not predictive of Offensive performance, but the kind of training was, as reported below.

<u>NTC Terrain Similarity Correlations.</u>

Since the distribution of terrain ratings was limited to five scores with the OPFOR included and four without, Spearman rank order correlations were calculated. The correlation between CXR and NTC terrain similarity on Offense with OPFOR included was .56 ($N = 15$, $p < .05$); and without the OPFOR, $r = .51$ ($N = 14$, $p < .05$). The correlation on Defense with OPFOR included was .49 ($N = 17$, $p < .05$), and without the OPFOR .41 ($N = 16$, $p < .06$).
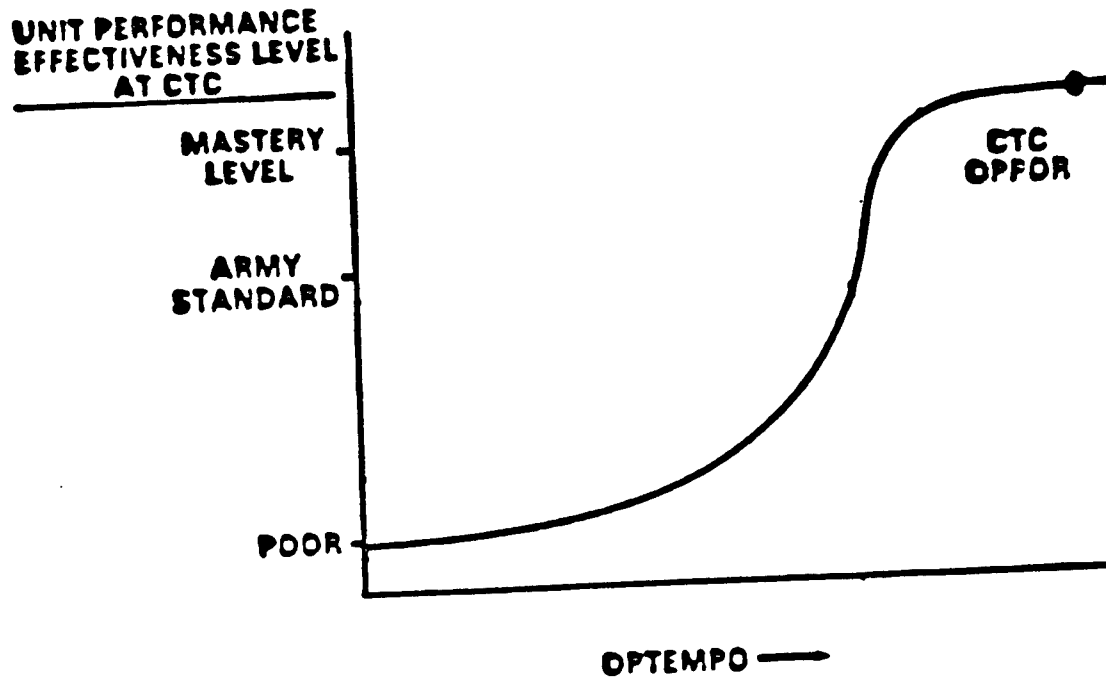
Figure 1. Hypothetical relationship of unit performance to OPTEMPO.
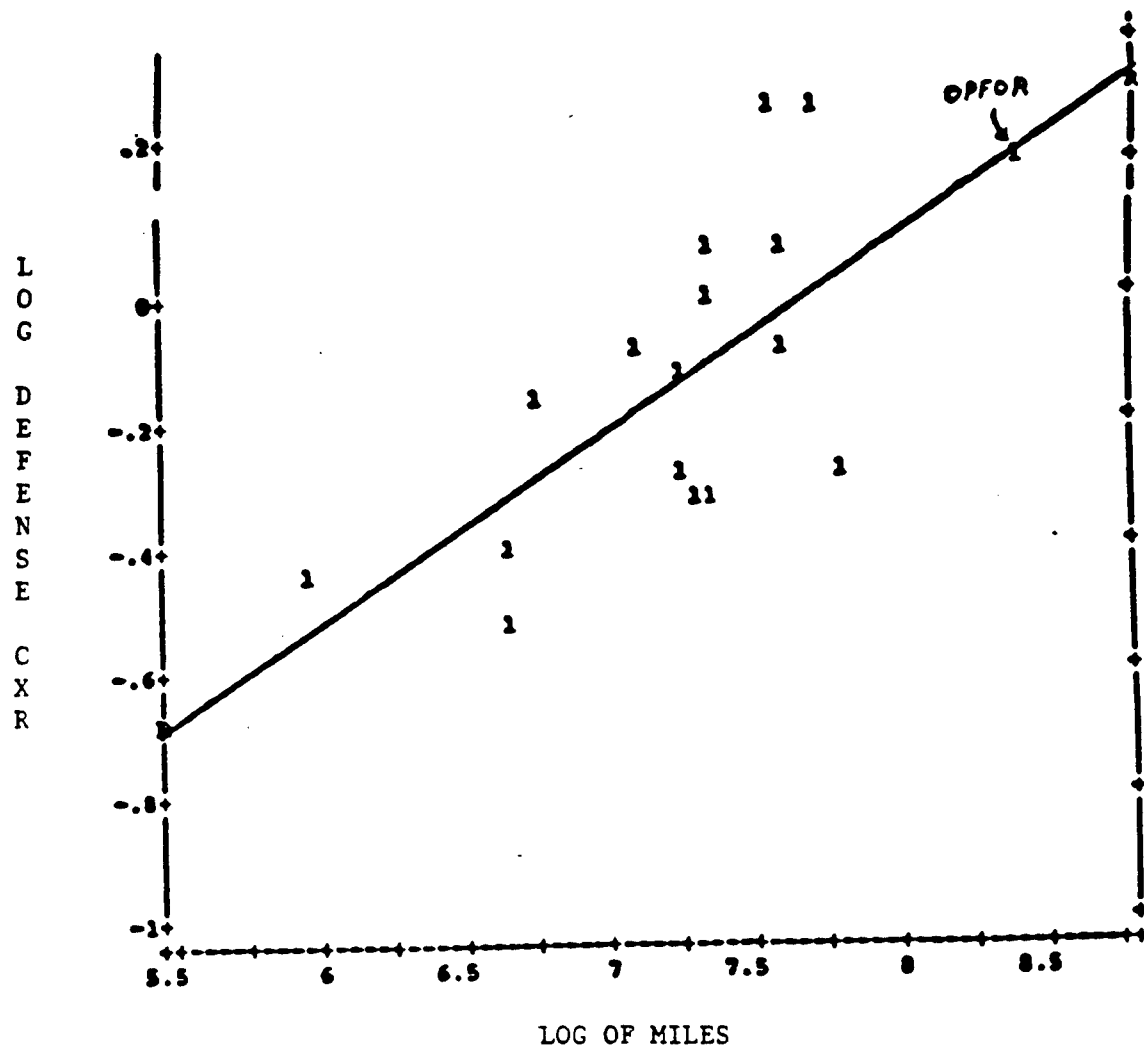
FIGURE 2. REGRESSION OF LOG DEFENSIVE CXR ON LOG TANK MILEAGE

Multiple Linear Regression Analyses.

The two predictors, tank mileage and NTC terrain similarity, were included in a set of stepwise multiple linear regression analyses, with and without the OPFOR, for Defensive and Offensive missions. However, on Defense, only tank mileage was selected, and on Offense, only NTC terrain similarity was selected at the .05 level of significance. The fact that the two predictors were not able to combine to yield a significantly improved prediction of performance was due to their own correlation, .67 with the OPFOR included (N = 17, p < .002) and .52 without the OPFOR (N = 16, p < .05); this correlation between NTC terrain similarity and training mileage (r = .52) implies a tendency for units with greater ranges to take advantage of that opportunity in training.

DISCUSSION

Objectively measured unit performance was found to be positively related to tank training mileage, or Ground OPTEMPO, on defensive missions; on average, as OPTEMPO increases, so does unit performance. Conduct of defensive operations gives an edge to the defender in knowing and organizing the defense to take advantage of terrain held, and that advantage afforded to the BLUFOR tended to neutralize the keen knowledge of terrain possessed by the OPFOR, thus allowing the level of BLUFOR combat proficiency to be revealed.

Although OPTEMPO was not correlated with Offensive mission performance, the similarity of home-station training areas to the NTC was. The OPFOR's vast experience gives them a great ability to anticipate where and how the BLUFOR will typically attempt to move and attack, and the OPFOR's complete mastery of the NTC terrain often enables them to confuse, slow, and decisively block the BLUFOR. Large scale combat operations such as those simulated at the NTC are extremely complicated to conduct; they involve combat support and combat service support operations, spread across thousands of meters, as well as the operations of the combat arms units whose performance was directly measured. Results on offense imply that training for offense requires the opportunity to practice movement and command, control and communication in large areas at the battalion task force level, as is possible at some posts but not others. The overall pattern of results for offensive and defensive missions demonstrates the benefit for amount of relevant practice, as pronounced by Thorndike's Law of Exercise (Thorndike[7]).

Another interesting training feature demonstrated by the results may be seen in Figure 2, where the performance variability on Defense associated with low mileage is relatively low, but the variability

178

associated with high mileage is high. It appears that units which have a very limited opportunity to train, necessarily perform poorly; whereas only some units with a large opportunity to train reap full benefits. Given the moderate size of the correlation between unit Defensive performance and NTC terrain similarity at home station (r = .41, or only about 16% of the performance variance is accounted for by terrain), the differences in unit performance with high OPTEMPO cannot be explained simply by differences in home-station training areas. Considering the unpredictable dynamics and complexity of the modern battlefield, the capability of leaders as trainers is also shown here to be important. In another ARI research project underway, variations in training preparation for the NTC are under study to determine the critical features of leadership, cohesion and training that lead to signifi- cant differences in performance capability as demonstrated at the NTC.

CONCLUSION

Results demonstrate that the methodology, devised to meet a requirement for determining if OPTEMPO relates to unit performance capability, is both practically feasible and useful. The results reported here have in fact been cited by the Secretary of Defense and used to defend the training budget.

References

1. Bryan, W.L. & Harter, N. Studies on the telegraphic language; the acquisition of a hierarchy of habits. Psychological Review, 1899,6, 345-375.

2. Carroll, J.B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

3. Cedel, T.E., and Fuchs, R.P. An analysis of factors affecting pilot proficiency. Air Force Center for Studies and Analyses, Washington, D.C. Final Study Report, DEC 1986.

4. DePoy, P.E. Observable effects of reduced flying hour programs. Center for Naval Analyses, Alexandria, VA, Memorandum for the Vice Chief of Naval Operations, (CNA) 84-1588, SEP 1984.

5. Hiller, J.H. Deriving useful lessons from combat simulations. Defense Management Journal, 1987, Second and Third Qtr., 28-33.

6. Horowitz, S. A., Hammon, C.P., and Palmer, P.R. Relating flying hour activity to the performance of aircrews: A progress report. Institute for Defense Analyses, Alexandria VA, DEC 1987.

7. Thorndike, E.L. The psychology of wants, interests and attitudes. New York: D. Appleton-Century, 1935.

A <u>Coxiella</u> <u>burnetii</u> Repetitive DNA Element with Potential Diagnostic and
Mutagenic Capabilities (U)

*Timothy A. Hoover, Dr.[1], Michael H. Vodkin, Dr.[2], Jim C. Williams, CDR[1,3]

[1]U.S. Army Medical Research Institute of Infectious Diseases, Intracellular
Pathogens Branch, Bacteriology Division, Fort Detrick, Frederick, Maryland
21701-5011

[2]University of Illinois, College of Veterinary Medicine, Department of
Pathobiology, Urbana, Illinois 61801

[3]Office of the Director of Intramural Research Programs, National Institute
of Allergy and Infectious Diseases, Bethesda, Maryland 20892

INTRODUCTION

Q (query) fever is an endemic and epidemic airborne and vector-borne
bacterial disease with worldwide distribution[1-6]. The disease is a result
of infection of host cells by the obligate intraphagolysosomal bacterial
pathogen, <u>Coxiella</u> <u>burnetii</u>[7]. <u>C</u>. <u>burnetii</u> is resistant to drying and
desiccation and is stable in various habitats that are normally
bactericidal for most non-sporulating bacteria[8]. Transmission of <u>C</u>.
<u>burnetii</u> to man and animals occurs by various routes, including aerosol
exposure, the bite of a tick, and ingestion of contaminated meats and dairy
products. Man is highly susceptible to Q fever, with only one
microorganism required to produce infection[9]. Seroepidemiology has shown
that the incidence of Q fever in humans exposed naturally ranges between 5%
and 30% in endemic and high-risk areas, respectively[10]. Occupational
exposure of humans to large animals, especially sheep, goats, and cattle,
is a clear risk factor[11-13]. Seroprevalence of Q fever in sheep flocks,
goat herds, cattle herds, stray dogs, and impounded cats within the United
States and Canada ranges between 7% and 24% positives, depending on the
seasonal variation [reviewed in 11]. In outbreaks of Q fever among
researchers studying sheep, up to 92% of the animals have been
seropositive[12]. The infection is associated with high morbidity and low
mortality in both animals and humans. The disease in animals ranges from
subclinical to abortion epidemics. Humans are highly susceptible to
infection by <u>C</u>. <u>burnetii</u>, with only one microorganism required to produce
clinical disease[9]. Q fever in humans is characterized by fever (100%),
chills (62%), headache (80%), myalgia (59%), arthralgia (22%), retrobulbar

pain (13%), chest pain (13%), cough (43%), nausea (30%), vomiting (30%), diarrhea (9%), hepatomegaly (42%), splenomegaly (29%), jaundice (3%), rash (5%), and pneumonia (51%)[14]. Liver function abnormalities include pathological effects in bilirubin (9%), alkaline phosphatase (34%) and aspartate aminotransferase (70%) levels[14]. Estimates, derived from primary cases of chronic Q fever, indicate that roughly 6% of infected individuals may develop chronic disease[15]. In cases of endocarditis, the aortic and mitral valves are usually colonized by C. burnetii.

Infection by C. burnetii induces both humoral[16] and cellular immune responses[17-20]. The cellular immune response is critical for the clearance of the microorganism from the affected animal[21,22]. Long-term cellular immunity has been noted in individuals suffering acute Q fever[17]. However, individuals with chronic Q fever may develop specific anergy[18-20]. Vaccinated individuals may develop both humoral and cellular immunity[23,24]. Humans appear to develop life-long protection after being infected, however, the protection afforded by vaccination is still in question.

The study of the genetics of C. burnetii complements the ongoing endeavors to obtain an efficacious vaccine, diagnostic reagents, and nucleic acid probes for early detection systems. Genetic analysis of C. burnetii is limited to the cloning of discrete segments of the genome into Escherichia coli and the restriction mapping of the chromosome and its 36 kilo-base pair (kbp) plasmid by means of agarose gel electrophoresis and, more recently, pulsed-field gel electrophoresis. The genes for citrate synthase[25], aspartate transcarbamoylase[26], a lipoprotein[27], and a 56 kilodalton (kd) heat-shock protein homologous to several bacterial common antigens[28] have been cloned and sequenced. An autonomous replicating sequence (ARS) likely to be the origin of replication has also been cloned and sequenced[29]. Comparison of restriction maps and Southern blots of virulent C. burnetii phase I and avirulent phase II revealed a large, 18 kbp deletion in phase II[30]. Phenotypic differences between the strains are related to altered lipopolysaccharide (LPS) and protein profiles[31,32]. Verification of the involvement of the deleted DNA region in LPS synthesis and in pathogenicity awaits DNA sequence analysis and restoration of genes discovered on this fragment to phase II cells, with concomitant conversion of the strain from avirulence to virulence.

A means by which manipulations of bacterial genomes are achieved involves the use of insertion sequences (IS) and transposons. These segments of DNA are moved enzymatically from one point on the chromosome to another frequently, often with a copy left at the first site. The insertion of DNA occurs randomly into the chromosome, often interrupting coding sequences and thus causing truncated, often non-functional proteins to be synthesized. This report discusses a region of C. burnetii DNA that is found in 15 to 20 copies on the chromosome and has structural features similar to typical insertion sequences. Such an element, endogenous to C. burnetii, provides a potentially powerful mutagenic and diagnostic tool. Its presence in multiple copies on the chromosome provides additional loci for homologous recombination between the chromosome and plasmids containing

HOOVER, VODKIN, WILLIAMS

the IS. IS elements with inserts of foreign DNA can carry the inserts onto
the chromosome during the integration event. Furthermore, DNA probes can
be made from this sequence and tested for uniqueness to C. burnetii.
Unless this IS is found in other bacteria, it may well be diagnostic of C.
burnetii, and may possibly be used to distinguish various C. burnetii
strains.

BACTERIAL INSERTION SEQUENCES[33-37]

IS elements are found in a wide variety of bacteria. The property in
common to all IS elements is the ability to transpose, or insert, into new
locations on the chromosome. These elements are from 150 to 1500 bp. The
ends are defined by short (10 to 50 bp) sequences that are nearly
identical, but are oriented inversely. The enzyme needed for the
transposition may be encoded within the IS element. The longer IS elements
contain genes for a transposase, but this is not a universal property. The
transposase recognizes and acts upon the inverted repeats, cutting the
double-stranded DNA at the ends of the repeats and causing these ends to
insert at random sites. Proteins that bind DNA generally have an alpha
helix-turn-alpha helix secondary structure motif, which fits in the double
helix groove of specific sequences. These proteins are often basic, with
an abundance of arginine, lysine, and histidine residues to counteract the
negatively charged DNA to which they bind. Often a short (3 to 20 bp)
duplication of the new site is found flanking the IS element on both sides,
indicating the likelihood of staggered cuts at the sites of insertion,
which are repaired by DNA polymerase fill-in. The length of the
duplication is characteristic for a given IS element. Transposons are
similar to IS elements but carry additional genes not required for
transposition, such as antibiotic-resistance genes. Transposons can be
flanked on both sides by entire IS elements, and the whole composite can
transpose as a unit.

COXIELLA BURNETII INSERTION SEQUENCE

The observation that a DNA segment carrying the htpAB (high temperature
inducible protein, or 56 kd common antigen) operon was present in multiple
copies on the C. burnetii chromosome suggested the presence of a repetitive
segment of DNA. When specific portions of the operon and flanking
sequences were used to make probes for Southern blots of restricted
chromosomal DNA, the htpAB operon was present in a single copy. However, a
region downstream from the operon was present in approximately 15 to 20
copies per chromosome. The plasmid carrying the htpAB operon and 2.5 kbp
of downstream flanking sequence was furthermore observed to be unstable.
Clones were occasionally isolated that were missing large fragments of the
parent plasmid. The 2.5 kbp segment was sequenced to help determine its
usefulness as a diagnostic probe, as multi-copy sequences are amplified

relative to other chromosomal genes. DNA sequence analysis revealed an IS-like element beginning 125 bp downstream of htpB (Fig. 1).

DNA sequencing was performed on an Applied Biosystems Model 370A DNA sequencer. This system uses primers labelled with fluorescent groups, as opposed to radioactive nucleotides, and Sanger dideoxy chain termination procedures[38]. Exonuclease III was used to generate deletions on the plasmid containing the IS element to obtain suitable overlapping templates for sequencing. Double-stranded sequencing still gives inconsistent results with this system. The best results are obtained with M13 single-stranded templates. The IS element, however, could not be cloned into M13 in its entirety, nor could portions that left the second ORF intact and downstream of the lac promoter found on M13mp18 and M13mp19. This suggested that the second ORF has transposase activity, or at least is capable of cutting the remaining inverted repeat in the set of deletions that removed the first repeat, thus making these constructs extremely unstable in M13. M13 goes through single-and double-stranded forms in its life cycle[39]. Transposase may be expressed from the double-stranded form of the M13 clone, which can be present in up to 200 copies per cell[39]. If ORF1 encoded a transposition inhibitor, which is often seen in IS elements to prevent excessive and deleterious insertions into the chromosome, its absence would further amplify the activity of the transposase, increasing the frequency of transposition events. Inverted repeats in the double-stranded form of the clone could serve as a substrate for transposase, leading to deletions and M13 mini-phage. These smaller forms quickly took over the phage population, making the yield of full-length forms too low for detection and sequencing. Carefully constructed subclones eventually allowed the sequence to be obtained from both strands.

The first feature of the sequence suggesting that this was an IS element was the presence of perfect 12 bp inverted repeats 1329 bp apart. Missing was the duplication flanking the inverted repeats. However, several recently described IS elements are also missing duplications[40-42]. The second feature characteristic of IS elements is the presence of two open reading frames (ORF) which would encode 163 and 161 amino acid polypeptides. These are small compared to other known transposases and ORFs of IS elements, but are larger than the two found in IS1 (91 and 124 residues). ORF1 has a predicted isoelectric point of 10.77, and that of ORF2 is 10.81, both very basic proteins. ORF2 also has a predicted helix-turn-helix structure from residues 118 to 138[43,44] (Fig. 2).

Two other random clones were selected from a C. burnetii genomic library by virtue of the presence of an unusually long 990 bp DdeI fragment (DdeI restricts DNA into fragments averaging about 250 bp). This fragment occurs within the htpAB IS element, and thus should be present in genomic subclones containing other copies of the IS. Restriction mapping verified that these clones contained the same restriction fragments as did the htpAB associated IS element, but, as was expected for IS elements at different loci, were missing sites outside the inverted repeat. Initial sequence analysis on one of these clones (double-stranded) showed that the inverted

HOOVER, VODKIN, WILLIAMS

Figure 1. Depiction of the htpAB operon and downstream region which includes an IS-like element. The scale below the line indicates base pairs in thousands. Open boxes represent known coding sequences or open reading frames. Arrows above the boxes show the direction of transcription. Filled box before htpA represents the heat-shock promoter. Small arrows below ORF1&2 show the location of the inverted repeats. There is a transcriptional terminator structure between htpB and ORF2.

HOOVER, VODKIN, WILLIAMS
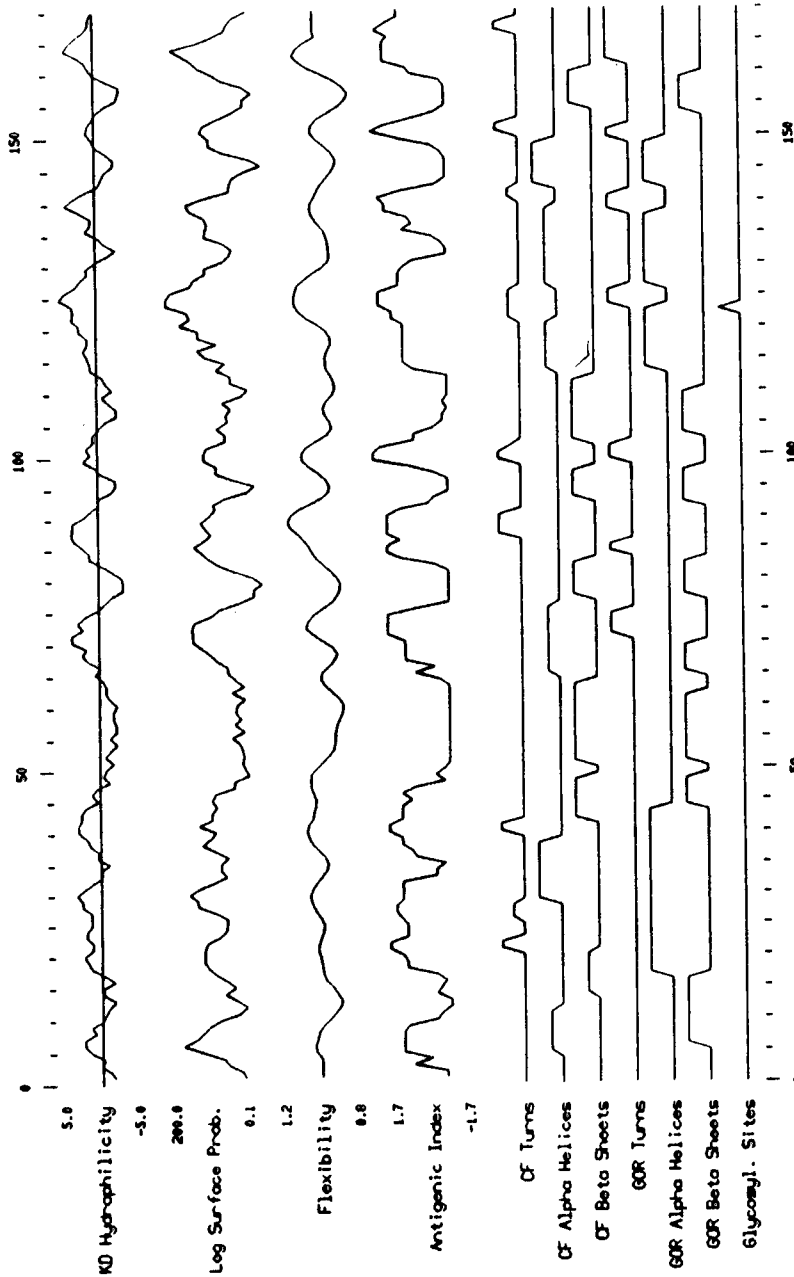
Figure 2. Graphic of secondary structure predictions for ORF2. Note the alpha helix-turn-alpha helix structure between residues 118 and 138, predicted by both CF (Chou and Fasman) and GOR (Robson and Garnier) algorithms. Program source: PeptideStructure and PlotStructure from the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin.

HOOVER, VODKIN, WILLIAMS

repeat upstream from ORF1 was identical to the htpAB IS repeat as were the first 50 or so internal bases. The region beyond the repeat was not the same as that in the htpAB IS, indicating that it is in a different chromosomal location. More careful DNA sequence analysis of the complete IS elements using single stranded templates should determine the extent of identity between the three elements and whether duplications are found at the two other insert locations.

DISCUSSION

Known IS elements from various bacterial species share little homology with one another[33]. The inverted repeats that define the ends of IS elements are likewise dissimilar. Therefore, transposases must also be different in their DNA-binding domains in order to bind specifically to the varied inverted repeats, and should show a corresponding lack of homology in the genes which encode them. This makes the IS an ideal candidate for a diagnostic probe for C. burnetii, in that a possibly unique sequence is amplified on the chromosome. IS elements show variable chromosomal locations, even within strains of the same species, so it may be possible to identify C. burnetii strain types via Southern blots of genomic DNA from cells recovered from infected hosts.

The discovery of an insertion sequence indigenous to C. burnetii may provide an extremely useful tool for experiments designed to determine genes required for virulence. Its presence in up to 20 copies on the chromosome suggests that the IS-encoded transposase is under less stringent regulation than most other IS elements, which are usually present in less than ten copies. These sites of insertion can be used as loci for homologous recombination to obtain stable integration onto the chromosome of practically any genes or DNA sequences desired. As those C. burnetii genes involved in pathogenicity and virulence are identified through mutational analysis, wild-type copies can be returned to the chromosome, and the restoration of virulence confirmed. Mutations in the IS itself may yield an altered form with even higher transposase activity, and may be useful in introducing mutations around the chromosome.

The ecological niche of C. burnetii makes it an excellent candidate for a universal vaccine vector. This microorganism is readily phagocytized by macrophages[45]. Growth of the microorganisms occurs in the phagolysosomal compartment[7], a vesicle that results from the fusion of a phagosome with lysosome, an acidic compartment. The usual fate of internalized bacteria is degradation within the phagolysosome, resulting in pH-dependent antigen processing[46,47]. For C. burnetii, this pH-dependent processing pathway triggers the microorganism to grow[7]. Proteins expressed by the microorganism within the endosomal, prelysosomal, and lysosomal compartments are potential substrates for the antigen-processing pathway[46,47]. An attenuated C. burnetii, destined to remain functional for only a short time, may be able to synthesize enough foreign antigen to induce a specific immune response, thus inducing protective immunity to the

pathogen. A shuttle vector may be adequate for sufficient expression of foreign antigens. In the event that the shuttle vector and the endogenous plasmid are incompatible, an alternative genetic mechanism for stabilizing the expression of a particular gene could rely on integration of that gene into the chromosome. The integration event is more likely to occur with the presence on the chromosome of multiple copies of a target sequence such as the IS.

The usefulness of C. burnetii as a live-attenuated vaccine vector derives from the stability of this microorganism under adverse environmental conditions. A cold chain and large amounts of inactivated antigen would not be necessary components for vaccines. Expression and presentation of protective proteins from various pathogens to the human and animal cells active in antigen processing would be facilitated.

The cloning of an ARS strengthens the possibilities of constructing a shuttle vector containing E. coli and C. burnetii origins of replication. The large endogenous C. burnetii Nine Mile I plasmid has also been cloned in its entirety in a cosmid vector. The origin of replication of the plasmid may also prove useful for constructing a shuttle vector. The requirement of other genes and the potential incompatibility of two origins on different plasmids in the same cell will be studied. The apparent low copy number of the endogenous plasmid in C. burnetii may alleviate somewhat the competition for stable maintenance and allow the plasmid and the shuttle vector to coexist long enough to allow expression of foreign proteins.

In addition to an origin of replication, a shuttle vector should have cloning sites for insertion of foreign DNA downstream from a functional transcription promoter. The 56-kd common antigen is expressed from a heat-shock promoter[28] that could prove useful in shuttle vector construction. This promoter is homologous to the heat-shock promoter in E. coli, and, in fact, causes an overproduction of the cloned C. burnetii common antigen in E. coli hosts grown at 42°C. Inclusion of this promoter in a shuttle vector would, therefore, allow testing for expression of foreign genes in E. coli prior to transforming the C. burnetii carrier strain.

A selectable marker, such as an antibiotic resistance gene, would be essential for selecting the appropriate shuttle vector construct in E. coli. Metabolic deficiencies have been used for other live bacterial vaccines,[48] such that only those cells with the wild-type gene introduced on a plasmid are capable of growth on selective media. The IS may be useful in creating C. burnetii mutants that only grow in the presence of a shuttle vector-borne gene and, by its presence on the same plasmid, foreign antigen genes.

SUMMARY

A region of the C. burnetii chromosome, present in 15 to 20 copies per cell, was sequenced and found to resemble typical bacterial insertion sequences (IS). The IS was 1353 bp, including two perfect 12 bp inverted

repeats. Two ORFs between the inverted repeats potentially encode 163 and 161 amino acid polypeptides; both are extremely basic. ORF2 had secondary structures similar to DNA binding proteins and transposases. ORF1 may encode a transposase inhibitor. The IS should facilitate the introduction of mutations or the insertion of genes of interest into the C. burnetii chromosome via homologous recombination.

REFERENCES

1. Babudieri, B. 1959. Q fever: a zoonosis. Adv. Vet. Sci. 5:81-182.
2. Berge, T. O., E. H. Lennette. 1953. World distribution of Q fever: human, animal and arthropod infection. Am. J. Hyg. 57: 125-143.
3. Kaplan, M. M., and P. Bertangna. 1955. The geographical distribution of Q fever. Bull. W. H. O. 13:829-860.
4. Tigertt, W. D. 1959. Studies on Q fever in man, pp. 39-46. In: Smadel, J. E. (ed.), Symposium on Q fever, Medical science publication no. 6, U. S. Government Printing Office, Washington, D. C.
5. Stoker, M. B. P., and B. P. Marmion. 1955. The spread of Q fever from animals to man. The natural history of a rickettsial disease. Bull. W. H. O. 13:781-806.
6. Wentworth, B. B. 1955. Historical review of the literature on Q fever. Bacteriol. Rev. 19:129-149.
7. Hackstadt, T., and J. C. Williams. 1981. Biochemical stratagem for obligate parasitism of eukaryotic cells by Coxiella burnetii. Proc. Natl. Acad. Sci., USA. 78:3240-3244.
8. Scott, G. H., T. F. McCaul., and J. C. Williams. 1989. Radiation inactivation of Coxiella burnetii. J. Gen. Microbiol. 135:3263-3270.
9. Tigertt, W. D., A. S. Benenson, and W. S. Gochenour. 1961. Airborne Q fever. Bacteriol. Rev. 25:285-293.
10. Marrie, J. J. 1990. Epidemiology of Q fever. ppxxx-xxx. In: Marrie, T. J. (ed.), Q fever the disease, Vol. I. CRC Press, Inc., Boca Raton, Fl. (In preparation).
11. Marrie, J. J., J. Van Buren, J. Fraser, E. V. Haldane, R. S. Faulkner, J. C. Williams, and C. Kwan. 1985. Seroepidemiology of Q fever among domestic animals in Nova Scotia. Am. J. Pub. Health. 75:763-766.
12. Simor, A. E., J. L. Brunton, I. E. Salit, H. Vellend, L. Ford-Jones, and L. P. Spence. 1974. Q fever: hazard from sheep used in research. Can. Med. Assoc. J. 130:1013-1016.
13. Marrie, T. J., and J. Fraser. 1985. Prevalence of antibodies to Coxiella burnetii among veterinarians and slaughterhouse workers in Nova Scotia. Can. Vet. J. 26:181-184.
14. Swayer, L. A., D. B. Fishbein, and J. E. McDade. Q fever: current concepts. Rev. Infect. Dis. 9:935-946.
15. Tellez, A., C. Sainz, C. Echevarria, S. Carlos (de), M. V. Fernandez, P. Leon, and R. Brezina. 1988. Q fever in Spain: acute and chronic cases, 1981-1985. Rev. Infect. Dis. 10:198-202.

HOOVER, VODKIN, WILLIAMS

16. Peacock, M. G., R. N. Philip, J. C. Williams, and R. S. Faulkner. 1983. Serological evaluation of Q fever in humans: enhanced phase I titers of immunoglobulins G and A are diagnostic for Q fever endocarditis. Infect. Immun. 41:1089-1098.

17. Jerrels, T. R., L. P. Mallavia, and D. J. Hinrichs. 1975. Detection of long-term cellular immunity to Coxiella burnetii as assayed by lymphocyte transformation. Infect. Immun. 11:280-286.

18. Koster, F. T., J. S. Goodwin, and J. C. Williams. 1984. Immunoregulation in acute chronic Q fever, pp. 282-284. In: Leive, L. and Schlessinger, D. (ed.), Microbiology - 1984. American Society for Microbiology, Washington, D. C.

19. Koster, F. T., J. S. Goodwin, and J. C. Williams. 1985. Selective antigen-specific anergy in patients with Q fever. J. Infect. Dis. 152:1283-1289.

20. Koster, F. T., J. C. Williams, and J. S. Goodwin. 1985. Cellular immunity in Q fever: modulation of responsiveness by a suppressor T cell-monocyte circuit. J. Immunol. 135:1067-1072.

21. Kishimoto, R. A., J. W. Johnson, R. H. Kenyon, M. S. Ascher, E. W. Larson, and C. E. Pedersen Jr. 1978. Cell-mediated immune responses of guinea pigs to an inactivated phase I Coxiella burnetii vaccine. Infect. Immun. 19:194-198.

22. Kishimoto, R. A., and J. S. Walker. 1976. Interaction between Coxiella burnetii and guinea pig peritoneal macrophages. Infect. Immun. 14:416-421.

23. Izzo, A. A., B. P. Marmion, and D. A. Worswick, D. A. 1988. Markers of cell-mediated immunity after vaccination with an inactivated, whole-cell Q fever vaccine. J. Infect. Dis. 157:781-789.

24. Ormsbee, R. A., and B. P. Marmion. 1990. Prevention of Coxiella burnetii infection: vaccines and guidelines for those at risk. ppxxx-xxx. In: Marrie, T. J. (ed.), Q fever the disease, Vol I. CRC Press, Inc., Boca Raton, Fl. (In preparation).

25. Heinzen, R. A., and L. P. Mallavia. 1987. Cloning and functional expression of the Coxiella burnetii citrate synthase gene in Escherichia coli. Infect. Immun. 55:848-855.

26. Hoover, T. A., and J. C. Williams. Characterization of Coxiella burnetii pyrB. Proceedings of the Eighth Sesqui-Annual Meeting of the American Society for Rickettsiology and Rickettsial Diseases. N.Y. Acad. Sci. In Press.

27. Hoover, T. A., and J. C. Williams. 1990. A Coxiella burnetii gene encodes a 29-kD antigen with a bacterial lipoprotein signal sequence, pp. 447-450. In: Vaccines 90. Brown, F., R. M. Chanock, H. S. Ginsburg, and R. A. Lerner (eds.). Cold Spring Harbor: Cold Spring Harbor Press.

28. Vodkin, M. H., and J. C. Williams. 1988. A heat shock operon in Coxiella burnetii produces a major antigen homologous to a protein in both mycobacteria and Escherichia coli. J. Bacteriol. 170:1227-1234.

29. Chen, S.H., T. Hoover, H. A. Thompson, and J. C. Williams. Characterization of an origin and DNA replication in Coxiella burnetii.

Proceedings of the Eighth Sesqui-Annual Meeting of the American Society for Rickettsiology and Rickettsial Diseases. N.Y. Acad. Sci. In Press.

30. Vodkin, M. H., and J. C. Williams. 1986. Overlapping deletion in two spontaneous phase variants of Coxiella burnetii. J. Gen. Microbiol. 132:2587-2594.

31. Amano, K., J. C. Williams, S. R. Missler, and V. N. Reinhold. 1987. Structure and biological relationships of Coxiella burnetii lipopolysaccharides. J. Biol. Chem. 262:4740-4747.

32. Williams, J. C., and S. Stewart. 1984. Identification of immunogenic proteins of Coxiella burnetii phase variants, pp. 257-262. In: Microbiology 1984, L. Leive and D. Schlessinger, (eds), American Society for Microbiology, Washington, D. C.

33. Syvanen, M. 1988. Bacterial insertion sequences, p. 331-356. In: Genetic Recombination, R. Kucherlapati and G. R. Smith (eds). American Society for Microbiology, Washington, D.C.

34. Iida, S., J. Meyer, and W. Arber. 1983. Prokaryotic IS elements, pp. 159-222. In: Mobile Genetic Elements, J. Shapiro, (ed), Academic Press, Inc., New York.

35. Grindley, N., and R. Reed. 1985. Transpositional recombination in procaryotes. Annu. Rev. Biochem. 54:863-896.

36. Craig, N., and N. Kleckner. 1987. Transposition and site-specific recombination, pp. 1054-1070. In: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, (eds). American Society for Microbiology, Washington, D.C.

37. Syvanen, M. 1984. The evolutionary implications of mobile genetic elements. Annu. Rev. Genet. 18:271-293.

38. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5463-5467.

39. Bethesda Research Laboratories, Inc. 1980. M13 Cloning/Sequencing Manual. Bethesda Research Laboratories, Gaithersburg, Maryland.

40. McPheat, W. L., J. H. Hanson, I. Livey, and J. S. Robertson. 1989. Analysis of the chromosomal location of two copies of a Bordetella pertussis insertion sequence. Mol. Microbiol. 3:985-989.

41. Barberis-Maino, L., B. Berger-Bachi, H. Weber, W. D. Beck and F. H. Kayser. 1987. IS431, a staphylococcal insertion sequence-like element related to IS26 from Proteus vulgaris. Gene 59:107-113.

42. McPheat, W. L., J. H. Hanson, I. Livey, and J. S. Robertson. 1989. Analysis of separate isolates of repeated DNA sequences from Bordetella pertussis. J. Gen. Microbiol. 135: 1515-1520.

43. Wolf, H., S. Modrow, M. Motz, V. A. Jameson, G. Hermann, and B. Fortsch. 1988. An integrated family of amino acid sequence analysis programs. CABIOS 4:187-191.

44. Jameson V. A., and H. Wolf. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. CABIOS 4:181-186.

45.   Williams, J. C., T. F. McCaul, H. A. Thompson, and D. M. Waag. 1989. Molecular strategies for uptake and phagolysosomal growth of Coxiella burnetii in nonimmune and immune hosts, pp. 127-140. In: Intracellular Parasitism, J. W. Moulder, (ed). CRC Press, Inc., Florida.

46.   Weaver, C. T., and E. R. Unanue. 1990. The costimulatory function of antigen-presenting cells. Immunology Today 11:49-55.

47.   Harding, C. V., Leyva-Corbian, F., and Unanue, E. R. 1988. Mechanisms of antigen processing. Immunol. Rev., 106:77.

48.   Dougan, G., L. Smith, and F. Heffron. 1989. Live bacterial vaccines and their application as carriers for foreign antigens, pp. 271-300. In: Vaccine Biotechnology, J. L. Bittle, and F. L. Murphy, (eds). Academic Press, Inc., California.

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

Evaluating Combat Soldier Energy Balance in the Field
with Stable Isotope and Ambulatory Monitoring Methods (U)

*Reed W. Hoyt, Ph.D., Tanya E. Jones, M.S., R.D.,
Harris R. Lieberman, Ph.D., LTC E. Wayne Askew, Ph.D.,
and Allen Cymerman, Ph.D.
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760-5007

INTRODUCTION

Is it normal for physically active soldiers in the field to be anorectic? Our research and the research of others suggests that soldiers in the field are in negative energy balance regardless of the type of ration consumed [1-11]. However, quantitative confirmation of these observations is lacking, primarily because of the technical difficulty of accurately measuring energy expenditure in the field with conventional methods. The doubly labeled water method, which we have recently validated for use in soldiers in the field [12], makes it possible to measure the energy expenditure of free-ranging humans without encumbering the soldier.

The purpose of this research was to meet the need for accurate energy balance measurements on free-ranging combat forces and to improve our understanding of how physically active soldiers allocate their time and energy resources. The doubly labeled water method, microprocessor-based ambulatory monitors, and standard nutritional techniques were respectively used to determine the total energy expenditures, average hourly rates of energy expenditure during activity, and the energy intake of combat forces in the field. Three squads of Marines engaged in a physically demanding 11-day cold weather field training exercise at 2600 m elevation were studied. The Marines consumed either the Improved Meal, Ready-To-Eat (IMRE), the Ration, Cold Weather (RCW), or the Ration, Lightweight (RLW).

METHODS

Subjects. Twenty-eight Marines, participating in the Winter Mountain Leader Course at the Marine Corps Mountain Warfare Training Center (MCMWTC), Bridgeport, CA, as either instructors or

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

students, volunteered for this 11-day study. The test volunteers
had been divided into three squads prior to the beginning of the
test. Each squad was randomly assigned to one of three
experimental rations: the Improved Meal, Ready-To-Eat (IMRE), the
Ration, Cold Weather (RCW), or the Ration, Lightweight (RLW). The
subjects gave their informed consent after being briefed on the
purpose, risks and benefits of the study. This study was approved
by the Human Use and Review Committee at the U.S. Army Research
Institute of Environmental Medicine, and by the U.S. Army Medical
Research and Development Command, Office of the Surgeon General.

Protocol. This experiment was conducted during the second two
weeks of a six-week Winter Mountain Leader Course at the MCMWTC.
The Winter Mountain Leader Course is designed to bring selected
Marines to a high standard of technical and tactical expertise in
the basic and advanced skills required for the successful conduct
of operations in a cold weather/mountainous environment. Students
are taught military skiing and snowshoeing, bivouac routines,
survival skills, avalanche avoidance and rescue, and mountain
navigation.

The MCMWTC training area encompasses 50,000 acres of semi-
arid, mountainous terrain and includes parts of the Toiyabe
National Forest. During the study, which was conducted in January,
ambient temperature ranged from -15 °C to +13°C (+5 to +55°F) and the
average snow depth was approximately 3 feet. On day 1 of the study
the Marines were transported by tracked vehicles from the MCMWTC
(2200 m/7218 ft elevation) to the Grouse Meadows training area
(2550 m/8353 ft). They established bivouac areas and engaged in
ski and snowshoe training. On days 2 and 3, the Marines relocated
bivouac areas, completed numerous day and night displacements on
skis and snowshoes, and stood sentry duty. On day 4, a 6 km ski
race and a night land navigation exercise were conducted. On day 5
the Marines skied 10 km back to the MCMWTC. During the weekend
(days 6 and 7), students and instructors were free to do as they
wished but agreed not to consume any food or beverages other than
that issued to them. On day 8, the Marines were deployed by
helicopter to Sardine Meadows, a second training area at 2210
m/7243 ft. Activities during days 8-11 included setting up bivouac
areas, participating in reconnaissance patrols, and a 10 km
biathalon. On day 11 the class returned to the MCMWTC by
helicopter.

Testing on days 0, 6, and 12 was performed at the MCMWTC. The
semi-nude body weights of overnight fasted subjects were measured
at approximately 0700 h on test days with a balance accurate to ±
0.01 kg. The schedule of heavy water administration, saliva and

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

urine sample collection, food log records, resting energy expenditure measurements, and activity monitoring is outlined in Table 1.

    Rations. There were eight subjects in the Improved Meal, Ready-To-Eat (IMRE) group; 10 subjects in the Ration, Cold Weather (RCW) group; and 10 subjects in the Ration, Lightweight (RLW) group. One subject from the RCW group was called back to regular duty on day eight and was therefore excused from the study. The IMRE group was issued four rations/man/day equal to approximately 5200 kcal/man/day (2.73 kg/day mass and 5.9 liter/day volume). The RCW group was issued one dehydrated ration/man/day which supplied approximately 4500 kcal/man/day (1.25 kg/day mass, 3.7 l/day volume). The RLW group was issued two dehydrated rations/man/day which supplied approximately 4000 kcal/man/day (0.86 kg/day mass; 1.2 l/day volume). Each subject was issued pocket-sized ration-specific logbooks in which to record ration consumption. Trained dietitians instructed the subjects individually how to accurately self-record food consumption in the log books. The subjects located each food component on a list of ration items and circled 1/4, 1/2, 3/4 or all of the food item that was consumed. On day 6 and at the end of the study the test subjects were interviewed by the same dietitians to ensure the accuracy and completeness of the recorded entries. Trading of ration components was only permitted within ration groups. Nutrient intakes were calculated by factoring individual food items consumed against known macro-nutrient values.

Table 1. Summary of experimental design.

DAY	˙0	1	2	3	4	5	˙6	˙7	8	9	10	11	˙12
DLW Dose	X												X
TBW (Saliva)	X												X
EE (Urine)	X	X				X							X
E Intake	X	X	X	X	X	X	X	X	X	X	X	X	
Activ. Mon.		X	X	X	X				X	X	X	X	
REE								X					
		<-EE Per.1->			<---------EE Per. 2----->								

    Key to abbreviations: ˙Days spent in garrison. DLW Dose, doubly labeled water dose; TBW (Saliva), saliva collection for total body water determination; EE (Urine), urine collection for energy expenditure determination; E Intake, self-recorded dietary intake; REE, resting energy expenditure; Activ. Mon., activity monitor data collected; EE Per. 1, energy expenditure period 1 (Days 1-4); EE Per. 2, energy expenditure period 2 (Days 5-11).

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

Doubly labeled water. Energy expenditure and changes in body energy stores were measured in a safe and unobtrusive manner with doubly labeled water ($^2H_2^{18}O$). Deuterium ($^2H_2$) and ^{18}O are safe, naturally occurring, nonradioactive isotopes. The subjects drank tracer doses of labeled water at the beginning of the time period of interest and then periodically provided urine and saliva samples. Changes in body energy stores were calculated from pre- and post-study measurements of total body water made by $H_2^{18}O$ dilution[13]. Energy expenditures were estimated from the difference in the rates of ^2H_2O and $H_2^{18}O$ elimination from the body. Deuterium leaves as water, mainly as urine, whereas ^{18}O leaves as both water and exhaled CO_2. The difference in the elimination rates is therefore equal to the rate of CO_2 production. A metabolic fuel quotient, i.e. a respiratory exchange ratio that reflected the final metabolic fuel mix combusted by the Marines, was calculated from macronutrient intakes and body fuel store utilization[14]. This permitted oxygen consumption and energy expenditure to be calculated from carbon dioxide production. The doubly labeled water method is particularly well suited for use in the field because the experimental procedures are safe, simple, and the amount of soldier time required is minimal.

Twenty-three subjects were administered doubly labeled water, while four subjects received unlabeled tap water. Changes in the base-line isotopic abundances in the group receiving no heavy water were used to correct the DLW calculations of energy expenditure in the experimental group. Each volunteer reported to the testing area on day 0 with a labeled screw top tube containing a base-line sample of their first morning void. They had refrained from eating or drinking anything for the previous 12 hours. Body weight was recorded. After base-line saliva samples had been collected, the subjects drank 0.30 g/kg body mass of $H_2^{18}O$ and 0.09 g/kg body mass of ^2H_2O, as well as the 100 ml of tap water used to rinse the container. Saliva samples were collected 3 and 4 hours after the doubly labeled water dose for total body water (TBW) determinations. Subjects were permitted to eat and drink following the final saliva collection. Samples of first morning void urines were periodically collected for energy expenditure measurements. On day 12 each experimental subject was given 0.10 g/kg body mass of $H_2^{18}O$, and 3 and 4 h post-dose saliva samples collected, for a final determination of total body water. The doses of heavy water administered were larger than typically used[16] due to the anticipated fluctuations in isotopic base-line and high rates of water intake. The urine and saliva samples were analyzed for ^{18}O and ^2H enrichment by Global Geochemistry, Inc. (Canoga Park, CA) using a Nuclide Isotope Ratio Mass Spectrometer (IRMS) (Nuclide, Inc., State College, PA).

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

Total body water was calculated by using ^{18}O enrichments in saliva before the dose, and at 3h and 4h after the dose using the formula[13]

$$TBW = (A/MW_d)(APE_d/100)\ 18.02\ [1/R_{std}(E_s - E_p)](1/1.01)$$

where A = dose in grams; MW_d = molecular weight of dose water; APE_d = atom percent excess enrichment of dose water; R_{std} = 2.005 x 10^{-3}, the ratio of heavy to light isotope of Standard Mean Ocean Water (SMOW); and E_s and E_p are the per mil (parts per thousand) enrichments of the final and pre-dose samples, respectively. The final division by 1.01 is necessary to adjust for the difference between actual TBW and the dilution space for ^{18}O.

The rate of CO_2 production was calculated using the equation[16]

$$rCO_2 = (N/2.078)(1.01k_o - 1.04k_H) - 0.0246rH_2O_f$$

where rCO_2 = rate of CO_2 production in mol; N = average of initial (day 0) and final (day 12) TBW measurements; k_o = fractional $H_2^{18}O$ turnover rate; k_H = fractional ^2H_2O turnover rate; and rH_2O_f = rate of fractionated evaporative water loss estimated as $1.05N(1.01k_o - 1.04k_H)$. The fractional turnover (isotopic elimination) rates for ^2H and ^{18}O were calculated over a given time period (t) by the two point method, which assumes single-compartment, steady-state kinetics, using the equation[16]

$$k = (\ln APE_f - \ln APE_i)/(t_2 - t_1)$$

where APE_i = initial enrichment, and APE_f = final enrichment, and $(t_2 - t_1)$ is the time elapsed between samples. The enrichments on day 1 were calculated relative to the individual pre-dose base-line isotopic abundances. Urine enrichments on days 5 and 12 were also adjusted for the corresponding change in the mean base-line isotopic abundances recorded in the group receiving no heavy water.

The metabolic fuel quotient was calculated from the macronutrient composition of the rations consumed and the amount of body fat and protein lost. Energy expenditure was then calculated from the rate of CO_2 production in liters ($\dot{V}CO_2$ l/min = rCO_2 mol/min · 22.4 l/mol) using conventional indirect calorimetric relationships[17].

Activity patterns. Activity patterns were recorded on days 1-4 and days 8-11 with compact (6.4x8.9x1.9 cm) lightweight (84 g) microprocessor-based monitors (Actigraph; Ambulatory Monitoring, Inc., Ardsley, New York) worn by the subjects. The activity

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

monitors transduce motor activity with a two-piece piezoelectric crystal sensitive to 0.01 g displacement force. Motor activity was quantified by recording the length of time the sensor output remained above the zero threshold during a continuous series of three minute epochs. These monitors, which were worn on the wrist and did not restrict motion or interfere with training duties, recorded and stored body movement data closely related to physical activity[15]. The monitors were not worn during days 5-7 while the Marines were in garrison. This permitted mid-experiment activity data retrieval.

Physical activity (A) was distinguished from inactivity with the equation $A = (-0.0003)X_{-2} + (-0.003)X_{-1} + (-0.011)X_0 + (-0.0003)X_{+1} + 1.004$. This equation is a modification of an algorithm for differentiating sleep from wakefulness[18]. The X_i's represent the activity monitor count for a completed three minute epoch. For example, X_{-2} is a measure for the three minute epoch completed 6 minutes ago. A given epoch (X_0) was scored as activity if $A < 0.5$, or scored as inactivity if $A \geq 0.5$. Activity data and data on total and resting rates of energy expenditure, were used to calculate the average rates of energy expenditure by the Marines during physical activity.

Resting gas exchange. On day 6, minute-to-minute gas exchange measurements were made on awake, postabsorptive, supine subjects using a face mask and an automated metabolic cart programmed for nutritional assessment (MMC Horizon System, Sensor Medics, Inc., Anaheim, CA). The resting energy expenditure and respiratory exchange ratio (RER) values were averaged over the final 3 minutes of each 20 minute measurement period.

Energy expenditure during inactivity and activity. Energy expended during physical inactivity was estimated by multiplying the rate of resting energy expenditure, estimated from the gas exchange measurements, by the amount of time the activity data indicated the Marine was inactive, and adding the specific dynamic action (SDA) or heat increment associated with the recorded ration consumption. The SDA was calculated as 6% of the total available dietary energy[19]. The mean rate of energy expenditure for a given period of physical activity was calculated by subtracting the energy expended during inactivity from total energy expenditure.

RESULTS

Physical characteristics of subjects. The physical characteristics and body composition of the subjects at the beginning of the study, and the changes in body composition that

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

occurred during the study, are shown in Table 2. There were no significant differences in physical characteristics among the three ration groups. All results are expressed as mean ± SEM.

Table 2. Physical Characteristics, body composition, and changes in body composition of the Marines.

	Age yr	Ht cm	Weight kg	TBW[1] kg	FFM[2] kg	FM[3] kg
Mean ± SE	27.2 0.8	178.3 1.4	79.6 1.4	44.2 0.7	60.2 1.0	19.4 0.8
Change ± SE	-- --	-- --	-2.5 0.3	-0.1 0.9	-0.1 0.3	-2.4 0.3

[1]TBW calculated from ^{18}O dilution space/1.01
[2]FFM = TBW/0.73
[3]FM = Body mass - Fat free mass

Ration macronutrient intakes. A detailed report of ration composition and consumption has been published[3]. The IMRE group consumed 62% of the 5192 kcal/man/day issued; the RCW group consumed 65% of the 4470 kcal/man/day issued; and the RLW group consumed 76% of the 4219 kcal/man/day issued. The average daily macronutrient intakes for each ration group are shown in Table 3. There were no significant differences in daily calorie or carbohydrate intakes among the three ration groups. The differences in protein consumption (IMRE>RLW>RCW) and fat consumption (RLW>IMRE>RCW) among the groups were significant.

Doubly labeled water energy expenditure. Base-line deuterium abundances decreased from -69 ± 5 ‰ on day 0, to -88 ± 4 ‰ on day 5, to -102 ± 4 ‰ on day 12. Base-line ^{18}O abundances were less pronounced, decreasing from -6.7 ± 0.6 ‰ on day 0, to -6.9 ± 0.8 ‰ on day 4, to -7.6 ± 0.6 ‰ on day 12. Isotopic elimination rates are shown in Table 4.

The metabolic fuel quotient, which reflects both food intake and body energy store combustion, was 0.84 ± 0.01 in the IMRE group, 0.87 ± 0.01 in the RCW group, and 0.84 ± 0.01 in the RLW group. The metabolic fuel quotient was significantly higher in the RCW group than in either the IMRE or RLW groups; the IMRE and RLW groups did not differ significantly. Ration energy intake, body energy store utilization, and energy expenditure are shown in Table 5.

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

Table 3. Mean macronutrient intakes of the members of the Improved Meal Ready-to-Eat (IMRE), the Ration, Cold Weather (RCW), and the Ration, Lightweight (RLW) groups receiving doubly labeled water.

Group	Energy (kcal)	Protein (g)	CHO (g)	Fat (g)
IMRE	3305 ± 217	137 ± 8	373 ± 26	140 ± 10
RCW	3015 ± 200	98 ± 5	423 ± 31	97 ± 4
RLW	3120 ± 299	112 ± 9	334 ± 31	151 ± 17
Mean	3132 ± 248	114 ± 8	376 ± 31	131 ± 13

Table 4. Deuterium elimination rate constants (k_D), and ^{18}O elimination rate constants (k_O).

Day	k_D (day^{-1})	k_O (day^{-1})
1-4	0.0649 ± 0.0048	0.1151 ± 0.0051
5-11	0.0926 ± 0.0037	0.1210 ± 0.0048
1-11	0.0809 ± 0.0040	0.1155 ± 0.0050

Table 5. Ration energy intake (E Intake), body energy store utilization (ΔES)[1], and energy expenditure (EE).

Day	E Intake (kcal/d)	ΔES (kcal/d)	EE (kcal/d)
1-4	3042 ± 257	---	6975 ± 232
5-11	3183 ± 242	---	3559 ± 228
1-11	3132 ± 248	-1672 ± 213	4816 ± 191

[1]ΔES was estimated from total body water measurements made pre- and post-experiment; body fat/carbohydrate energy stores were assumed to have an energy density of 7700 kcal/kg.

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

Resting gas exchange. Resting $\dot{V}_{O_2}$ averaged 0.265 ± 0.008 $l \cdot min^{-1}$, or 1.26 ± 0.04 $kcal \cdot min^{-1}$, with no significant differences among the three ration groups. Resting RER was 0.73 ± 0.01 in the IMRE group, 0.80 ± 0.02 in the RCW group, and 0.76 ± 0.02 in the RLW group. The RCW group had a significantly higher RER than the IMRE group ($P<0.05$); there were no other significant between-group differences.

Activity patterns and average hourly rates of energy expenditure. The Marines were active 74.72 ± 0.93 % of the time (17.93 h/day; 1076 ± 13 min/day) on days 1-4. On days 8-11 the subjects were active 68.95 ± 1.00 % of the time (16.54 h/day; 992 ± 14 min/day). During days 1-4, the subjects were physically inactive an average of 366 ± 13 $min \cdot day^{-1}$ during which time they expended 469 ± 28 kcal. During days 1-4, energy expended in physical activity was 352 ± 13 kcal/h or 5.9 ± 0.2 kcal/min. Energy expenditure during activity did not differ significantly among the ration groups. The subjects did not wear activity monitors while in garrison (days 5-7), therefore average hourly rates of energy expenditure during activity could not be calculated for the second energy expenditure period covering days 5-11.

DISCUSSION

These results, and the findings from our previous study where the doubly labeled water method was used to measure the energy expenditure of free-ranging soldiers working at light to moderate intensities[12], quantitatively confirm research that indicates that physically active soldiers in the field are normally in negative energy balance regardless of the type of ration consumed (Table 6)(1-11). Why are soldiers engaged in field training exercises anorectic? The mechanism is obscure, although an abrupt switch from a palatable, freely chosen diet to standard military field rations explains part of the decreases food intake. For example, when soldiers in the field were served hot A rations at regularly scheduled meal times the apparent limit to solid food consumption of field rations was exceeded and caloric balance achieved[20].

Animals engaged in activities important to survival, such as territorial defense or migration, commonly reduce food intake and lose body mass, even when food is available. These animal anorexias are thought to have adaptive value[21,22]. It is difficult to ascertain the adaptive value of anorexia to soldiers operating in a field environment, however, it is conceivable that anorexia may help soldiers adapt to demanding situations by increasing the amount of time available for more important military tasks. Secondarily, a reduction in ration consumption would benefit the

Table 6. Energy and carbohydrate (CHO) intakes from Ration Lightweight (RLW); Meal Ready-to-Eat (MRE), and Ration Cold Weather (RCW) consumption, and estimated energy expenditure (EE) in recent ration tests.

Ref.	Ration	Subject No.	Days	EE (kcal)	Wt. Loss (kg)	Avail. kcal	Intake (kcal)	Intake (%)	CHO Intake
1	RLW	36	12	3150	-2.5	2000	1780	89	210
2	RLW	17	30	3275	-5.2	2000	1945	97	197
3	RLW	10	11	4500	-2.8	4220	3205	76	345
4	MRE III	15	12	3820	-2.8	ADLIB	2285	-	250
5	MRE VI	126	11	3535	-1.7	3670	2515	69	270
5	MRE VII	129	11	3950	-2.4	4015	2515	63	290
5	MRE VIII	117	11	4150	-2.2	3940	2840	72	375
2	MRE VI	17	30	3250	-1.8	3600	2782	77	318
6	MRE III	10	6	2000	0	ADLIB	1790	-	190
6	MRE III	8	6	3250	-1.0	ADLIB	2150	-	234
7	MRE III	16	10	3950	-3.0	4892	2733	56	302
3	MRE VIII	8	11	4500	-2.1	5200	3217	62	369
8	MRE VI	32	10	3460	-2.2	4816	2009	45	210
8	MRE VIII	31	10	3855	-1.6	4571	2802	62	325
9	MRE III	27	34	2905	-3.7	3600	2189	61	246
10	MRE VI	167	3	3030	-	3670	2445	67	254
7	RCW	18	10	3950	-2.6	4541	2751	61	384
3	RCW	10	11	4500	-2.3	4470	2892	64	410
11	RCW	30	8	4700	-2.8	4470	2880	64	385
Mean				3670	-2.4		2512		293
SEM				155	0.3		102		16

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

soldier by reducing the mass and volume of rations that would need to be carried.

Limited ration consumption and high rates of energy expenditure, as in the Marines, naturally leads to an increased reliance on body fat stores (Table 2). Humans and animals maintain large depots of energy dense triglycerides that have no immediate metabolic function, but serve solely as readily-mobilized energy reserves available to meet any shortfall in food energy intake[23]. For example, a typical young male soldier weighing 74 kg (163 lb) has approximately 13.5 kg (29.6 lbs) of body fat[24]. This is equivalent to 69,300 kcal, assuming a body fat energy density of 7700 kcal/kg (3500 kcal/lb) and that two-thirds or 9 kg (19.8 lbs) of this body fat can be used without encroaching on nerve sheath lipids or other fats necessary for normal physiological function. This fat reserve, which constitutes approximately 98% of the body's energy reserve[25], is enough energy to meet a 2000 kcal/day energy deficit for over a month.

The Marines in this study weighed approximately 80 kg with a somewhat higher 19.5 kg body fat content. This is equal to 100,000 kcal, assuming 13 kg is body fat reserves. Approximately one-fifth of this fat reserve was utilized during the 11 days of this study. This suggests that during prolonged, physically demanding field operations, soldiers with adequate body fat stores would have an advantage over unusually lean soldiers.

There are, however, potentially maladaptive aspects to semi-starvation. Carbohydrate reserves constitute only around 2% of the body's energy reserves, and consequently have a limited capacity to buffer shortfalls in dietary carbohydrate intake[25]. When dietary carbohydrate intake does not meet metabolic demand, limited carbohydrate reserves are depleted and there is a switch to a fat-predominant fuel metabolism. A fat-predominant metabolism is characterized by carbohydrate sparing for nerve cells, red blood cells and other tissues with an obligate requirement for carbohydrates, a loss of lean body mass[2], and a decrease in physical performance[25]. When fat is the primary metabolic fuel, there can be as much as a 50% reduction in endurance exercise capacity and exercise intensities greater than 55% of maximum aerobic capacity cannot be sustained[26].

The Marines exercised at an average intensity equal to 84% of the mean maximum rate of energy expenditure measured by portable respirometer in soldiers performing short duration combat type activities[27]. An adequate dietary intake of easily-combusted carbohydrates would improve a soldier's or Marine's ability to

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

sustain this high average exercise intensity. However, except for
the RCW group in this study, carbohydrate intakes by soldiers in
the field have never attained the 400 g/man/day minimum suggested
by the Committee on Military Nutrition of the National Research
Council[28](Tables 3,6).

Recent studies have demonstrated that solid or liquid
carbohydrate supplements can be used to improve the carbohydrate
intake of soldiers in the field[8,29], and lessen the physiological
consequences of a blunted appetite (anorexia)[30]. In addition, there
is increasing evidence that field rations should be designed
primarily to be palatable, easily consumed, and to meet the
metabolic demand for carbohydrates, and only secondarily attempt to
achieve overall fat and energy balance[28,30]. This approach will not
solve the complex problem of anorexia, but will help alleviate the
metabolic consequences of being in a negative energy balance.

In summary, (1) during physically demanding field exercises
soldiers are characteristically anorectic and in negative energy
balance, and, (2) field rations that are designed primarily to meet
the metabolic demand for carbohydrates can alleviate the
physiological consequences of limited ration consumption.

REFERENCES

1. Siegel, S.F., Poole, P.M., Askew, E.W., Kinney, M.A., Shaw,
   C., Aylward, J., and H. Susan. Twelve-day field test of
   ration, lightweight, 30-day at Fort Chaffee, Arkansas.
   Natick/TR-87/032, 1985.
2. Askew, E.W., Munro, I., Sharp, M.A., Siegel, S., Popper, I.R.,
   Rose, M., Hoyt, R.W., Martin, J.W., Reynolds, K., Lieberman,
   H.R., Engell, D., and C.P. Shaw. Nutritional status and
   physical and mental performance of special operations soldiers
   consuming the Ration, Lightweight or the Meal, Ready-to-Eat
   military field ration during a 30 day field training exercise.
   USARIEM Technical Report T7-87, 1987.
3. Morgan, T.E., Hodgess, L.A., Schilling, D., Hoyt, R.W., Iwanyk,
   E.J., Mcaninch, G., Wells, T.C., Hubbard, V.S., and E.W. Askew.
   A comparison of the Meal, Ready-to-Eat, Ration, Cold Weather
   and Ration, Lightweight nutrient intakes during moderate
   altitude cold weather field training operations. USARIEM
   Technical Report T5-89, 1988.
4. Askew, E.W., Claybaugh, J.R., Cucinell, S.A., Young, A.J., E.G.
   Szeto. Nutrient intakes and work performance of soldiers
   during seven days of exercise at 7,200 feet altitude consuming
   the Meal, Ready-to-Eat ration. USARIEM Technical Report T3-87,
   1986.

5. Popper, R. Hirsh, E., Lesher, L., Engell, D., Jezior, B., Bell, B., and W.T. Matthew. Field evaluation of improved MRE, MRE VII, and MRE IV. Natick/TR-87/027, 1987.

6. Askew, E.W., Claybaugh, J.R., Hshiro, G.M., Stokes, W.S., Sato, A., and S.A. Cucinell. Mauna Kea III: metabolic effects of dietary carbohydrate supplementation during exercise at 4100 m altitude. USARIEM Technical Report T12-87, 1987.

7. Roberts, D.E., Askew, E.W., Rose, M.S., Sharp, M.A., Bruttig, S., and J.C. Buchbinder. Nutritional and hydration status of special forces soldiers consuming the Ration, Cold Weather or the Meal, Ready-to-Eat ration during a ten day cold weather field training exercise. USARIEM Technical Report T8-87, 1987.

8. Edwards, J.S.A., Roberts, D.E., Morgan, T.E., and L.S. Lester. An evaluation of the nutritional intake and acceptability of the Meal, Ready-to-Eat consumed with and without a supplemental pack in a cold environment. USARIEM Technical Report T18-89, 1989.

9. Hirsch, E., Meiselman, H.L., Popper, R.D., Smits, G., Jezior, B., Lichton, I., Wenkam, N., Burt, J., Fox, M., McNutt, S., Thiele, M.N., and O. Dirige. The effects of prolonged feeding Meal, Ready-to-Eat (MRE) operational rations. Natick/TR-85/035, 1983.

10. USACDEC, Fort Ord, CA and USARIEM, Natick, MA. Combat field feeding system force development test and experimentation (CFFS-FDTE). Test Report CDEC-TR-85-006A, 1986.

11. Roberts, D.E., McGuire, B.J., Engell, D.B., Salter, C.A., and M.S. Rose. The role of water consumption on consumption of the Ration, Cold Weather. USARIEM Technical Report T13-89, 1989.

12. Delany, J.P., Schoeller, D.A., Hoyt, R.W., Askew, E.W., and M.A. Sharp. Field use of $D_2^{18}O$ to measure energy expenditure of soldiers at different energy intakes. J. Appl. Physiol. 67:1922-1929, 1989.

13. Schoeller, D.A., Van Santen, E., D.W. Peterson, Deitz, W., Jaspan, J., and P.D. Klein. Total body water measurement in humans with ^{18}O- and ^{2}H-labeled water. Am. J. Clin. Nutr. 33:2686-2693, 1980.

14. Schoeller, D.A., Ravussin, E., Shutz, Y., Acheson, K.J., Baertschi, P., E. Jequier. Energy expenditure by doubly labeled water: validation in humans and proposed calculations. Am. J. Physiol. 250:R823-R830, 1986.

15. LaPorte, R.E., Kuller, L.H., Kupfer, D.J., McPartland, R.J., Matthews, G., and C. Caspersen. An objective measure of physical activity for epidemiological research. Am. J. Epidemiol. 109:158-168, 1979.

16. Schoeller, D.A. Measurement of energy expenditure in free-living humans by using doubly labeled water. J. Nutr. 118:1278-1289, 1988.

HOYT, JONES, LIEBERMAN, ASKEW, CYMERMAN

17. Lusk, G. The Elements of the Science of Nutrition (4th ed.). New York: Academic, 1976, p. 65.
18. Cole, R.J., and D.F. Kripke. Progress in automatic sleep/wake scoring by wrist actigraph. Association of Professional Sleep Societies, San Diego, CA, 1988, (Abstr.).
19. Nutrition and the Adult. Macronutrients, R. Alfin-Slater and D. Kritchevsky, eds, Plenum Press, New York, 1980, p. 80.
20. Rose, M.S., and D.E. Carlson. Effects of A-Ration meals on body weight during sustained field operations. USARIEM Technical Report T2-87, 1986.
21. Mrosovsky, N., and D.F. Sherry. Animal anorexias. Science 207:837-842, 1980.
22. King, J.R., M.E. Murphy. Periods of nutritional stress in annual cycles of endotherms: fact or fiction? Amer. Zool. 25:955-964, 1985.
23. Symposium on Lipids in Animal Life Histories. Amer. Zool. 16, 1976.
24. Fitzgerald, P.I., Vogel, J.A., Daniels, W.L., Dziados, J.E., Teves, M.A., Mello, R.P., and P.J. Reich. The body composition project: A summary report and descriptive data. USARIEM Technical Report T5/87, 1986.
25. Sahlin, K. Metabolic changes limiting muscle performance. In: Biochemistry of Exercise VI. International Series on Sports Sciences 16:323-344, 1986.
26. Costill, D.L. Carbohydrates for exercise: dietary demands for optimal performance. Int. J. Sports Med. 9:1-18, 1988.
27. Goldman, R.F. Energy expenditure of soldiers performing combat type activities. Ergonomics 8:321-327, 1965.
28. Military Nutrition Research Annual Report on Calorie Dense Rations, 1988. Food and Nutrition Board, National Research Council, Washington, D.C. 20418.
29. Jones, T.E, Hoyt, R.W., Baker, C.J., Hintlian, C.B, Walczak, P.S., Shaw, C.P., Schilling, D., Askew, E.W., and A. Cymerman. Voluntary consumption of a liquid carbohydrate supplement by special operations forces during a high altitude cold weather field training exercise. USARIEM Technical Report (In preparation.) 1990.
30. Military Nutrition Research Annual Report, 1989. Food and Nutrition Board, National Research Council. National Academy Press: Washington, D.C. 20418 (In press).

Solvent Stress Cracking and Failure Mechanisms in Polyetherimide Composites (U)

Alex J. Hsieh, Ph.D.
U.S. Army Materials Technology Laboratory
Polymer Research Branch, SLCMT-EMP
Watertown, MA 02172-0001

INTRODUCTION

During the past decade, polymeric composites have found increased use as structural materials in many military applications. However, there has also been increased concern with their environmental stability. For example, delamination crack growth has been observed in helicopter rotor blades as a result of fatigue loading. The subsequent diffusion, accumulation, and condensation of moisture can induce matrix microcracks, and cause components to fail under stresses well below their ultimate strengths.

Delamination between plies is considered one of the predominant modes of damage in service. Composites based on thermoplastic matrices have shown improvements in the interlaminar strengths over thermosets in composites (1). However, some of these thermoplastics are susceptible to chemical attack. The Army Materials Technology Laboratory (MTL) has been working on evaluating the durability of polymeric composites and transparent polymers in various service environments including chemical warfare simulants and a decontaminant, DS2. This is partly in response to the Nuclear, Biological,and Chemical (NBC) Contamination Survivability, Army Materiel Policy AR 70-71, which requires that all Army system materiel be hardened against degradation in a NBC environment. This hardening includes materiel survivability as well as the ability to be decontaminated.

Decontamination will subject materiel to organic solvents and surface active agents raising the question whether these liquids will compromise composite reliability and life expectancy. This subject has received little attention in the literature or in the Army assessment.

Previous work includes an analysis of the degradation mechanism of polycarbonate (PC) by DS2 (2), and an evaluation of Parylene C as a chemical resistant coating for PC (3). A study of the effect of physical orientation on the chemical resistance of polymethyl methacrylate (PMMA)

was also carried out (4). Results indicated that none of these
transparent polymers were free from the attack of most organic media.

This paper addresses studies utilizing polyetherimide to investigate
solvent induced failure mechanisms in neat resin and in composites
including adhesively bonded composites. Calculations of residual thermal
strains, which can drive cracks, and a model for delamination crack
growth are also presented.

MATERIALS AND EXPERIMENTAL METHODS

The polyetherimide, Ultem 1000, neat resin plaques, obtained from
General Electric Co., were 12.7 mm thick. Two kinds of polyetherimide
composites with different PEI matrix formulations were used in this work.
CYPAC 7010 and CYPAC 7005 composite prepregs, with respective Ultem 1000
and Ultem 5001 matrices, were obtained from American Cyanamid. CYPAC
7010 has AS-4 graphite woven fabric (3K70) in plain weave, and CYPAC 7005
consists of glass woven fabric (7781) with 8-Harness Satin weave pattern.
Fiber volume content is 67% in both materials. Ultem 1000 film in
various thicknesses were obtained from General Electric.

Experiments of solvent enhanced stress cracking were carried out via
fracture mechanics based tests. The compact tension (CT) specimen per
ASTM E399 (5), shown in Figure 1, was used to measure crack propagation
in the neat resin. The double cantilever beam (DCB) specimen, shown in
Figure 2, was used to study Mode I delamination crack growth in the
composites. In the latter, the sheets of prepreg were laid up into
12-ply composite plates and compression molded. For the adhesively
bonded composite specimen, an additional Ultem 1000 film adhesive was
embedded between the middle plies. During layup, a 0.075 mm thick layer
of Teflon-coated glass fabric was inserted between the center plies at
one end. This was to provide a starter notch for subsequent crack
propagation. Laminate fabrication was completed following the
manufacturer's suggestion at 315 °C under 0.7 MPa pressure. Hinges were
bonded onto both sides of the specimens at the end adjacent to the
starter notch, as shown in Figure 2, using film adhesive FM 123 from
American Cyanamid.

Specimens were tested under constant load with a static dead-weight
loading apparatus. Solvent effects were determined by dropwise addition
directly to the crack tip. Crack propagation measurements as a function
of time were made with either a traveling microscope or a camera. Since
these measurements were made on two different specimen geometries, the
crack growth data obtained for both neat resin and composite specimens
were needed to be reduced in terms of a common parameter in order for
comparison. The parameter used to measure the resistance to crack
propagation is called the strain energy release rate, G_I. Subsequently,
all the raw data of crack length versus time were reduced in terms of
crack growth rate, da/dt, versus G_I, for both neat resin and composite
specimens.

208

Fracture surfaces of the test specimens were coated with gold-palladium and examined using a Jeol-840 scanning electron microscope (SEM).
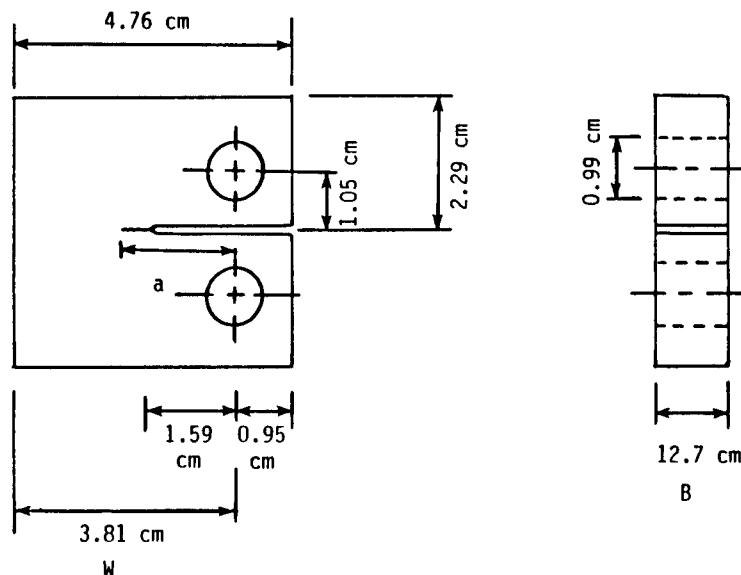
Figure 1
Compact tension
specimen geometry
per ASTM E399-83
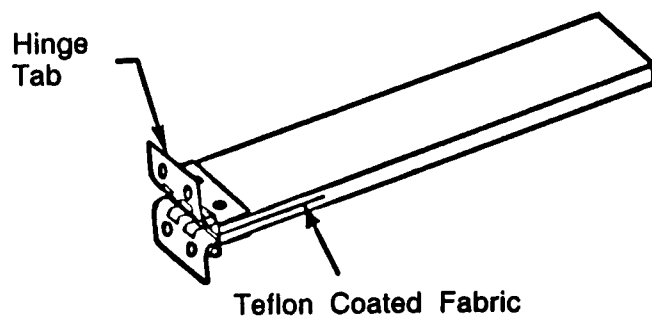
Figure 2
Sketch of double
cantilever beam
specimen

RESULTS AND DISCUSSION

Solvent Stress Cracking of Neat Resin

Crack growth in o-xylene for the Ultem 1000 neat resin occurred at G_I values, shown in Figure 3, which were much lower than the G_{Ic} value (3.2 kJ/m^2) needed to grow a crack in air (6). This indicated that o-xylene significantly reduced the resistance to crack propagation for the Ultem 1000 neat resin.

Figure 4 shows a SEM micrograph of the fracture surface with a discontinuous growth band pattern for the Ultem 1000 neat resin specimen

tested in o-xylene. These band markings oriented parallel to the advancing crack front are similar to those under fatigue crack growth in air for many glassy amorphous polymers (6).
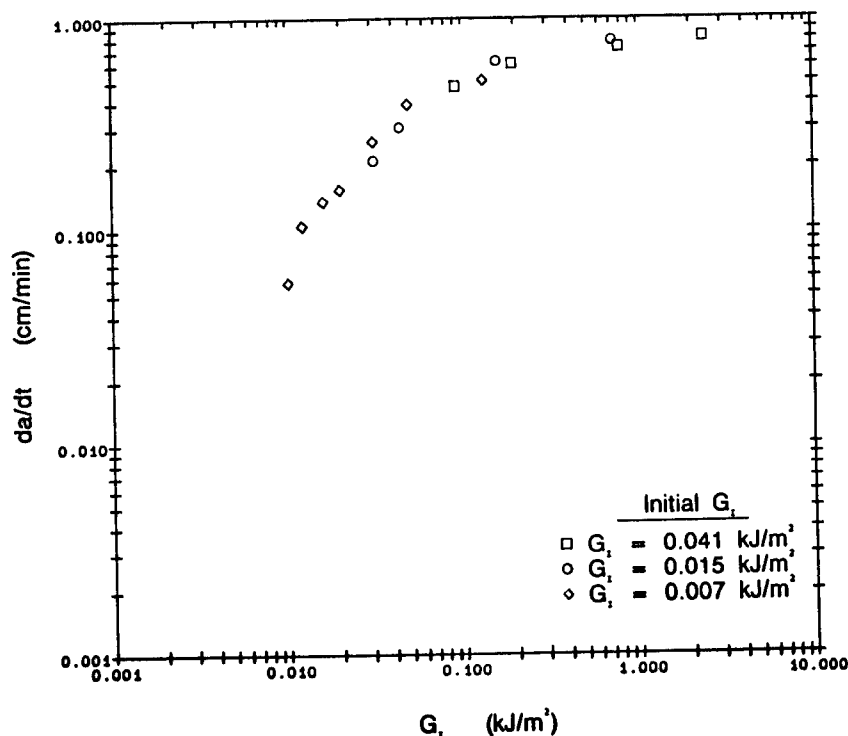
Figure 3
Crack growth rate versus G_I for the Ultem 1000 neat resin tested in o-xylene at various initial G_I values
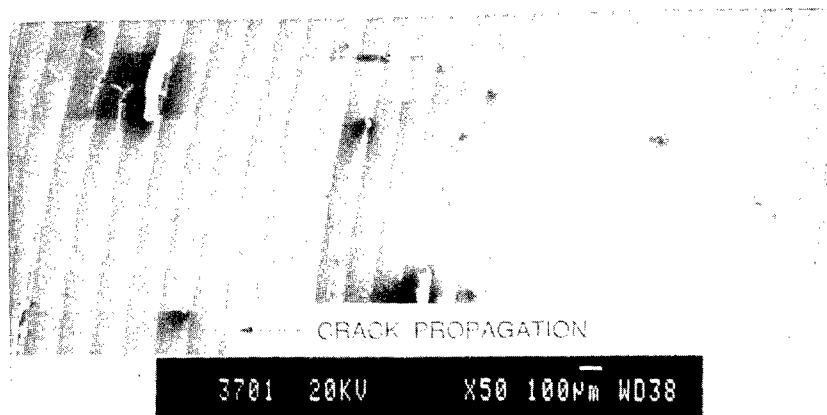
Figure 4
SEM micrograph of the fracture surface shows a discontinuous growth band pattern

Solvent Stress Cracking of Composites

Results in Figure 5 show that crack propagation in o-xylene for composites with Ultem 1000 matrix occurred at very low G_I levels compared to the critical value in air of G_{IC} = 2.2 kJ/m^2 (6). This indicates that

solvent would cause a significant reduction in composite delamination resistance as in the neat resin. However, the growth rates in the composites were much slower than those in the neat resin at most G_I values when compared in Figure 6. This is contrary to the lower G_{Ic} value in air for the composite versus the neat resin (2.2 vs. 3.2 kJ/m^2).

A fractographic study was also carried out on the composite specimens. For the specimens tested in o-xylene, there were matrix microcracks oriented normal to the delamination plane seen at the weave crossover regions in Figure 7. This was the result of a resin-rich zone at the crossovers of the woven fabric, and poor resistance to o-xylene of the Ultem 1000 matrix. On the contrary, ductile deformation of the matrix was seen at the crossover regions when these specimens were tested in air, as shown in Figure 8.

As described in more detail later, the presence of the matrix crack pattern on the specimens tested in o-xylene is attributed to residual thermal stress driven solvent stress cracks. The matrix stress was inherently produced from the cool down process due to the mismatch in thermal expansion coefficients between the constituents as well as a high glass transition temperature. However, the residual thermal stresses alone, without o-xylene, do not create matrix cracks in air.
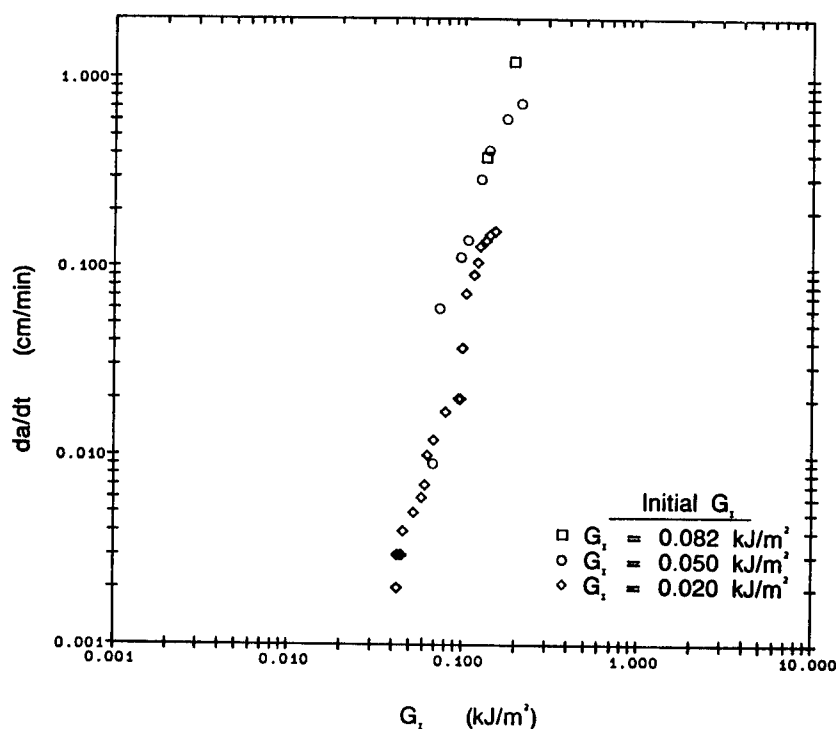
Figure 5 Crack growth rate vs. G_I for the Ultem 1000 composites tested in o-xylene at various initial G_I values
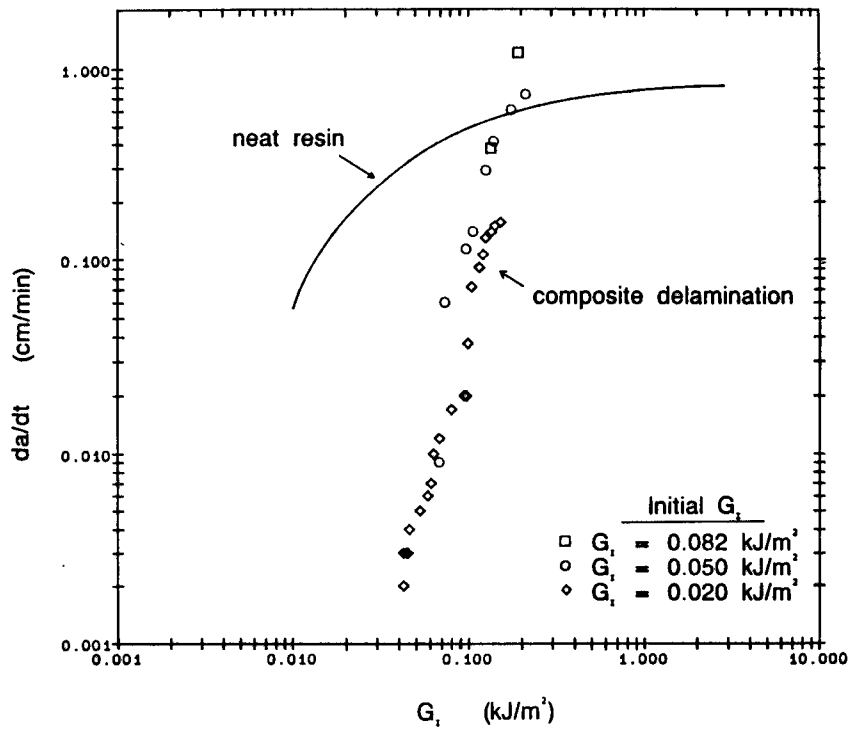
Figure 6
Comparison of
the Ultem 1000
composite data
with the data
for the Ultem
1000 neat resin
(solid line from
Figure 3)

Figure 7
SEM micrograph
shows matrix
microcracks at
the crossover
resin rich
regions for the
Ultem 1000
composite tested
in o-xylene
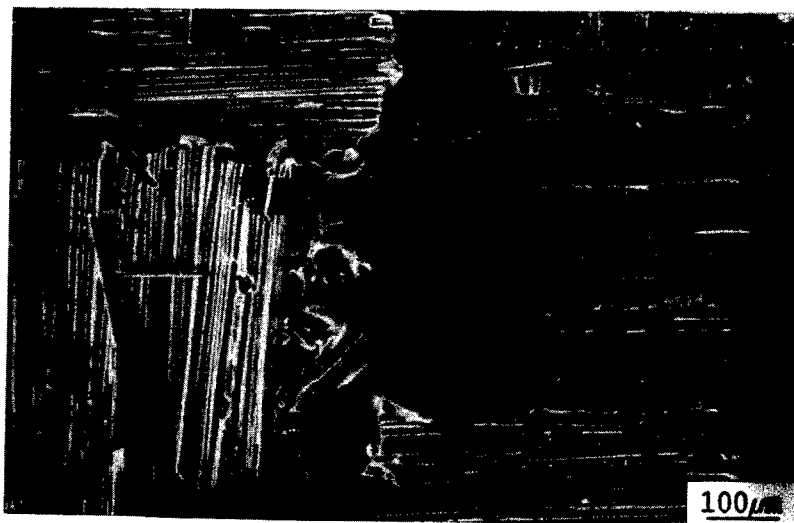
Figure 8
SEM micrograph
shows ductile
deformation of
the matrix
material at the
crossover resin
rich regions for
the Ultem 1000
composite tested
in air

Solvent Stress Cracking of Adhesively Bonded Composites

The importance of the residual stress on solvent stress cracking was further elaborated by testing composites with a central adhesive layer of Ultem 1000 film. Delamination in composites was expected to be similar to interlaminar crack growth in adhesively bonded composites because both processes involve the separation of materials held together by the polymer resin. Moreover, there would be always a residual thermal stress present in the adhesive/matrix layer. The stress is due to the mismatch of thermal expansion coefficients between the film adhesive and the composite, as well as a high glass transition temperature of the film adhesive.

A set of test specimens was prepared containing Ultem 1000 film in various thicknesses incorporated into an Ultem 5001 matrix based composite. Since the Ultem 5001 matrix in the composite substrate had better resistance to o-xylene than the Ultem 1000 adhesive matrix, crack propagation in o-xylene for these adhesively bonded specimens occurred primarily through the adhesive layer. In addition, cracks were seen perpendicular to the adhesive plane on the edges of the specimen and ahead of the main crack. The spacing of the cracks increases with an increase in film thickness (6). When the main crack was propagated through the adhesive, the resulting fracture surfaces showed a characteristic mosaic pattern of matrix cracks normal to the adhesive plane. A typical mosaic crack pattern is shown in Figure 9 for the adhesively bonded composite specimen with 0.46 mm thick adhesive. The thinner the adhesive layer, the closer the spacing of the mosaic cracks (6). In addition, the crack spacing seen at the specimen edges during

the test had been found to be consistent with the spacings of the mosaic cracks on the interior of the fracture surfaces.

Figure 9
SEM micrograph shows the mosaic crack pattern on the fracture surface of the composite specimen with 0.46 mm thick adhesive tested in o-xylene

Results indicated that the residual tensile stress field was sufficient to drive the solvent assisted stress cracks through the adhesive prior to the growth of the main crack. When the main crack passed, it needed to continually reinitiate failures in the individual blocks of remaining adhesive, which were still well bonded to the laminate on each side. Ultimately, the main crack propagated through the precracked adhesive in the form of coalesced small cracks in the mosaic blocks, and cohesive failure of the adhesive was the only mode seen in the specimens tested in o-xylene. The processes of crack initiation and crack propagation are sketched as shown in Figure 10. The nucleation point and crack growth band markings within a mosaic block are evident in Figure 11. Such a discontinuous growth band pattern was found on the fracture surfaces of the Ultem 1000 neat resin specimens tested in o-xylene.

Immersion Tests of Adhesively Bonded Composites

In order to verify the significance of the residual stress effect, some of the adhesively bonded composite specimens were tested in o-xylene immersion without any additional mechanical loading. These specimens were cut into a size of 25.4 mm by 25.4 mm. Results showed that adhesive cracks were clearly seen in the adhesive layer, including both edges and center-cut cross sections, for the immersion tested specimens. Figure 12 illustrates the presence of these cracks in the adhesive layer, in profile. Thus, the residual thermal stresses and strains are sufficient to drive the solvent stress crack system.
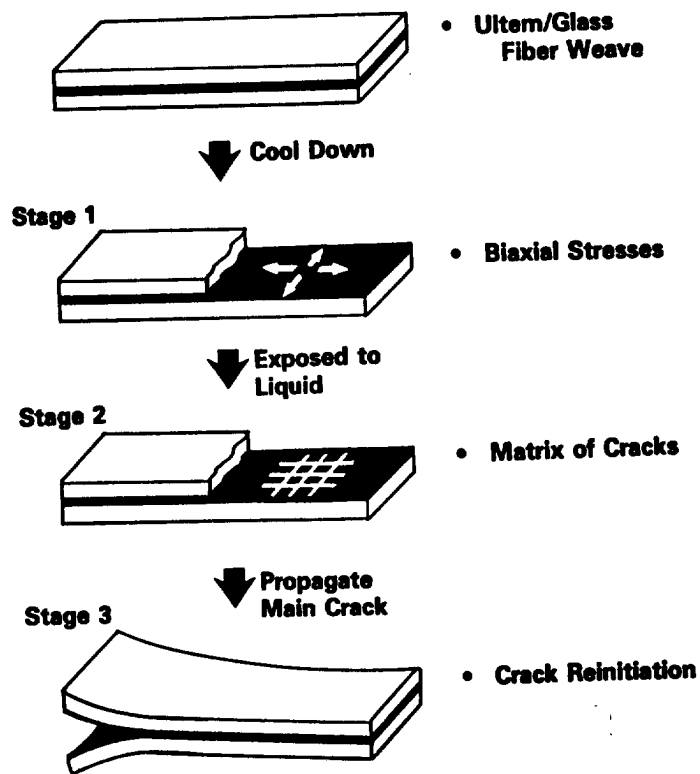
Figure 10
A sketch
illustrates the
processes of
crack initiation
and crack
propagation

Figure 11
SEM micrograph
shows nucleation
point and growth
band markings
within a mosaic
block

Immersion tests were also carried out in other solvents, such as toluene, 1,2,4-trimethyl benzene, hexanol, and isopropanol. Results of the specimen with 18 mil thick film are summarized in Table 1. Cracks were clearly seen in the adhesive layer when tested in toluene, and 1,2,4-trimethyl benzene, as in o-xylene. However, cracking did not occur when exposed to hexanol and isopropanol.

There is also a good correlation between the reported critical strain values for the Ultem 1000 neat resin and resistance to adhesive cracking in the test specimens. Values of the critical strain determined by Kambour et al (7) for the selected solvents are included in Table 1. Solvents associated with lower critical strain values promoted adhesive cracks without any additional mechanical loading. However, cracks did not occur in both hexanol and isopropanol, indicating that residual thermal strain in the adhesive layer was below the critical strain range of 1.26%.
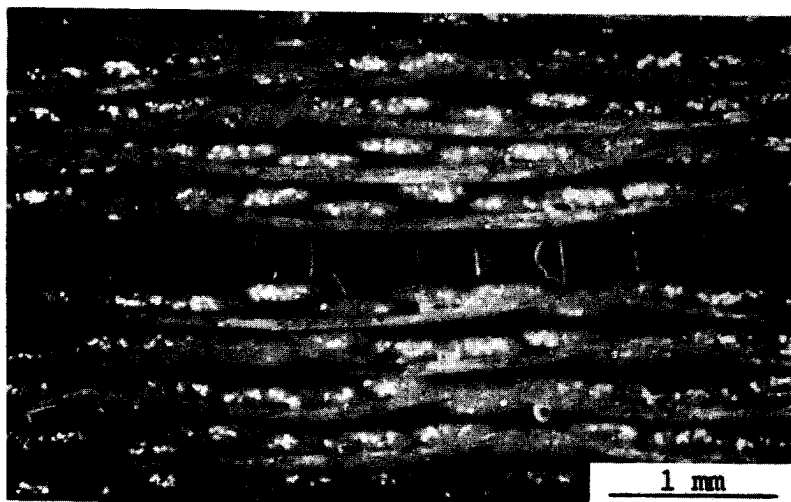
Figure 12
SEM micrograph
shows matrix
microcracks in
the adhesive
layer after the
o-xylene
immersion

Residual Thermal Strain Calculation

The extent of the residual thermal strain in the adhesive layer was calculated via a first approximation and by lamination theory. In the first approach the residual thermal strain, ε_T, was determined from difference in thermal expansion between the adhesive layer and the composite. That is,

$$\varepsilon_T = (\alpha_{adh} - \alpha_{comp}) (T_g - T_r) \qquad (1)$$

where α_{adh} and α_{comp} are coefficients of thermal expansion for adhesive and composite respectively, T_g is the glass transition temperature of the

adhesive, and T_r is room temperature. This equation assumes that thermal stress or strain develops in the adhesive layer when temperature is below T_g. This is because the molecular mobility is significantly reduced below T_g, and stress build-up can no longer be relaxed. Values of α are assumed to be constant over the temperature range, and material is assumed linear elastic.

α_{adh} and α_{comp} are 56 x 10^{-6}/°C and 8 x 10^{-6}/°C respectively (6), and T_g was determined as 216°C (6). From Eq. (1), ε_T was calculated to be 0.92%, which is close to residual strain values calculated based on lamination theory (6). These calculated strains are significantly higher than the critical strain, 0.5%, needed for cracking of Ultem 1000 neat resin in o-xylene as shown in Table 1. This implies that cracks can initiate in the adhesive layer simply due to the presence of residual thermal stress when these composite specimens are exposed to o-xylene. However, the higher critical strain of 1.26% required for stress crazing in hexanol and isopropanol indicates that the residual thermal strain is not sufficient to drive solvent crazes in the absence of additional mechanical loads for these solvents.

Calculation of G_I values for growth within the mosaic blocks

The growth of delamination cracks in adhesively bonded composites has been shown to occur primarily in the adhesive layer because of the poor resistance of Ultem 1000 to o-xylene. This results in a characteristic mosaic crack pattern. An approximate approach to calculate the G_I values needed for local cracks to grow within the mosaic block has been developed.

The sketch in Figure 13 illustrates that a uniform stress, σ, is assumed to be applied on a mosaic block which is ahead of the main crack. The local stress intensity factor associated with this stressed mosaic block can be regarded as the same K_I for a circular crack, with crack diameter equal d, embedded in a solid at the same stress, σ. K_I for the latter is given as (8)

$$K_{I,L} = 2\ \sigma\ (d/2\pi)^{1/2} \qquad (2)$$

where d is the embedded crack diameter, and $K_{I,L}$ is the local stress intensity factor within the mosaic block, with crack size equal to d.

The corresponding local strain energy release rate for growth, $G_{I,L}$, can be obtained as

$$G_{I,L} = (K_{I,L})^2/E_{adh}$$

$$= 4\ \sigma^2\ (d/2\pi)/E_{adh} \qquad (3)$$

where E_{adh} is Young's modulus of the adhesive. The stress, σ, in the above equation will be related to the apparent measured strain energy

release rate, $G_{I,APP}$, which is given as

$$G_{I,APP} = (P^2/2B)(dC/da) \qquad (4)$$

It is assumed that adhesive materials generally yield locally in air at the critical G_{Ic} with the corresponding yield stress, σ_{ys}. However, in o-xylene, crack propagation occurs at significantly lower G_I values. Therefore, the failure stress of the adhesive is reduced, and is assumed to vary proportionally to $(G_{I,APP}/G_{I,L})^{1/2}$. Thus,

$$\sigma = \sigma_{ys} (G_{I,APP}/G_{Ic})^{1/2} \qquad (5)$$

This reduced stress on the adhesive can be taken as the local stress needed for local crack growth within the block. Therefore, Eq. (3) can be rewritten as

$$G_{I,L} = 4 \sigma^2 (d/2\pi)/E_{adh}$$
$$= (2d/\pi) (\sigma_{ys})^2 (G_{I,APP}/G_{Ic})/E_{adh} \qquad (6)$$

Equation (6) allows calculation of the local G_I value for crack growth within any mosaic block with respect to the corresponding apparent macroscopic strain energy release rate, $G_{I,APP}$.
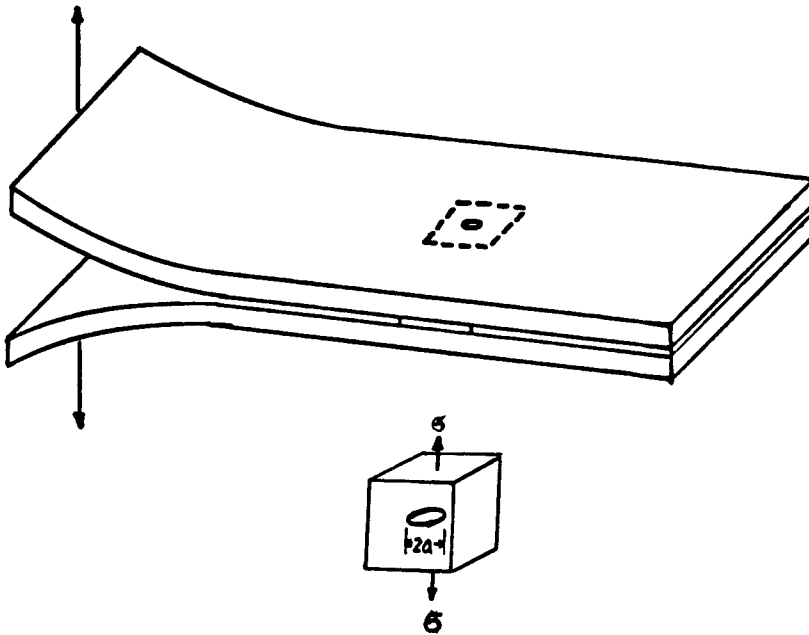
Figure 13
A sketch illustrates that a uniform stress, σ, is assumed to be applied on a mosaic crack which is ahead of the main crack

As an example, a mosaic block located in an adhesively bonded composite specimen with 0.152 mm thick adhesive film had a macroscopic G_I value determined to be 0.104 kJ/m^2 at a macroscopic crack velocity of about 0.9 mm/min (6). The crack diameter associated with this mosaic block was equivalent to 145 μm, measured from a SEM micrograph. Values of σ_{ys}, G_{Ic}, and E_{adh} were given as 1.05 x 10^8 N/m^2, 3.2 kJ/m^2, and 3 x 10^9 N/m^2 respectively. The local G_I value driving local mosaic block growth calculated via Eq. (6) is 0.011 kJ/m^2, which is almost an order of magnitude smaller than the measured apparent G_I value of the main crack, 0.104 kJ/m^2.

The shift in local vs. macroscopic G_I value is about an order of magnitude over most of the range. This is similar to the shift in the neat resin vs. composite or adhesive crack growth curves in Figure 14 (9). In particular, the threshold G_I values for solvent stress crack growth are about an order of magnitude lower for the neat resin, consistent with the prediction of this simple model. The reason for the slower main crack growth in the composites and adhesives is the lower local G_I associated with the very small cracks embedded within the mosaic blocks, which determine the macroscopic growth rate as they coalesce to form the main crack. The rate of crack growth in the adhesive bonded composite is comparable to that in the neat resin when the da/dt vs. G_I curve of the former is shifted to an order of magnitude smaller G_I values.
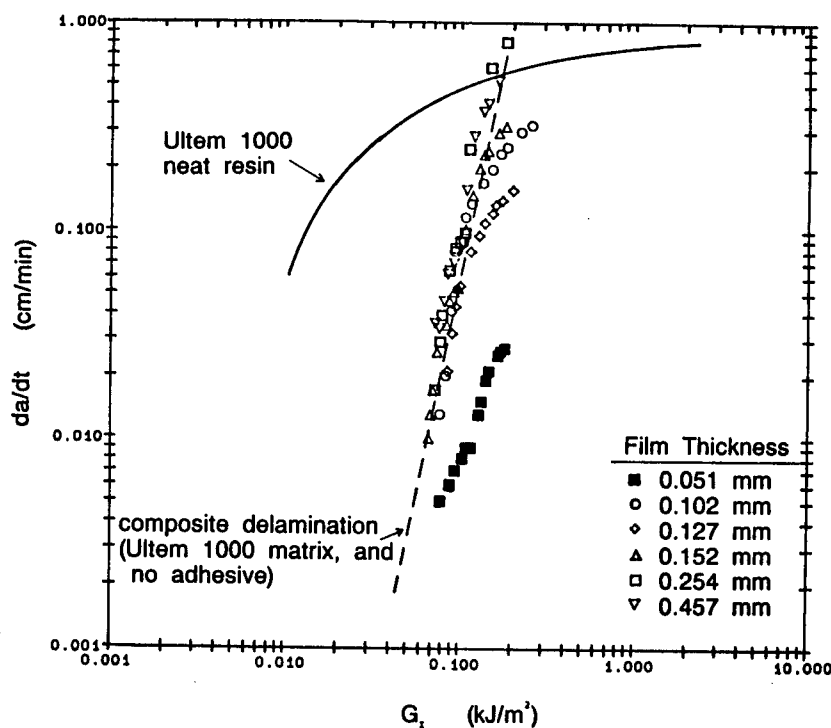
Figure 14 Comparison of the adhesively bonded composite data with the data for the Ultem 1000 neat resin (solid line from Figure 3) and Ultem 1000 composite delamination (dashed line from Figure 5)

The calculated applied G_I value for the delamination, $G_{I,APP}$ has no physical significance in cases where the actual cracks in question are small isolated cracks ahead of the main separation zone. The local G_I for each individual embedded crack is what governs growth of the local cracks, and, ultimately, growth of the main crack.

CONCLUSIONS

Results of this work show that exposure to solvent results in a significant increase in crack propagation in the Ultem 1000 resin based materials. Crack propagation in neat resin, composites, and adhesively bonded composites in o-xylene occurs at significantly lower G_I levels than the respective G_{IC} value needed to force crack growth in air. In this work, residual thermal stresses are high enough to drive solvent stress cracks both in the matrix and in the adhesive layer. These microcracks, even after solvent removal, will expose composite structures to the service conditions and NBC environment. Subsequently, fatigue loading, moisture uptake, and liquid absorption will cause composite deterioration and catastrophic failure. The findings of this study provide a basis for the assessment of composite durability in various environments.

REFERENCES

1. D. L. Hunston, R. J. Moulton, N. J. Johnston, W. D. Bascom, ASTM STP 937, 74 (1987).
2. L. H. Lee, J. J. Vanselow, MTL Technical Report TR 87-46, September (1987).
3. L. H. Lee, J. J. Vanselow, MTL TR 87-6, January (1987).
4. J. J. Vanselow, A. J. Hsieh, MTL TR 88-24, July (1988).
5. ASTM, Designation E399-83, 519 (1983).
6. A. J. Hsieh, PhD Thesis, Dept. of Materials Sci. & Eng., Massachusetts Institute of Technology, (1989).
7. S. A. White, S. R. Weissman, R. P. Kambour, J. Appl. Polym. Sci., 27, 2675-2682 (1982).
8. H. L. Ewalds, R. J. H. Wanhill, "Fracture Mechanics", Edward Arnold Publishers, 49 (1984).
9. A. J. Hsieh, J. F. Mandell, N. S. Schneider, accepted for publication in Polymer Composites.

Table 1

Results of immersion tests of adhesively bonded composites in various solvents

solvent	solubility[*][a] parameter	critical[a] strain,%		stress cracking
O-xylene	9.0	0.5	CK	Y
Toluene	8.9	0.58	CK	Y
1,2,4 Trimethyl Benzene	9.0	0.96	CK	Y
Hexanol	10.8	1.26	Z	N
Isopropanol	12.0	1.26	Z	N

*: $(cal/cm^3)^{1/2}$
a: data from Kambour et al., (Ref. 7)
CK: cracking agent
Z: crazing agent
Y: stress cracking occurred
N: stress cracking did not occur

Object Extraction for Automatic Target Recognition (U)

*John L. Johnson, Dr.
U.S. Army Missile Command
Redstone Arsenal, AL 35898-5248

Introduction

The problem of automatic target recognition from a missile-mounted video camera must account for many factors. The missile has six dynamic degrees of freedom, the perceived scene is two dimensional realtime video imagery transformed by the sensor, and the three dimensional scene contains numerous objects which may be in motion themselves and only some of which are the desired targets. Accordingly, the problem decomposes into two subproblems. The first is the issue of how to account for the dynamical variations of the perceived scene and the second is how to identify the targets. The latter problem can be solved by use of optical matched filter correlators,[1,2,3] but that application is dependent upon the solution of the former problem.[4] This paper will show how to describe the dynamical scene variations in such a way that the scene parameters of translation, rotation, range/scale and velocity changes can in principle be measured and used to generate mappings of these parameters. When these parameters are known, the perceived scene can be transformed into an invariant representation of the actual three dimensional scene which can in turn be used as an appropriate input to an optical matched filter correlator.

Analysis

A function f of space and time under an infinitesimal contact transformation[5] will have a new numerical value f′ related to the untransformed value by

$$\delta f = f' - f. \tag{1}$$

To first order, a Taylor's series expansion can be used to express this change in terms of the infinitesimal changes in the arguments of the function and its first partial derivatives with respect to these arguments. Suppose that each variable is transformed by the combined effects of infinitesimal changes in translation, rotation, scale, and velocity. If T, R, S, and L are the four corresponding transformation operators, then for small changes the sequence T*R*S*L gives

$$TRSL = (1+\delta T)(1+\delta R)(1+\delta S)(1+\delta L)$$
$$\approx 1 + \delta T + \delta R + \delta S + \delta L.$$

Any other sequence will obviously yield the same first-order result. The effect of each type of transformation can accordingly be evaluated separately and used to determine the combined effect in this first order approximation.

a) Translation
   Let a prime denote the transformed variables. Then

$$x' = x + dx \tag{2a}$$
$$y' = y + dy \tag{2b}$$
$$z' = z + dz \tag{2c}$$
$$t' = t + dt \tag{2d}$$

b) Rotation

To define the infinitesimal rotation construct a vector $\vec{d\theta}$ whose direction is the axis of rotation and whose magnitude is the amount of right-handed rotation. The new variables are then related to the old ones by

$$x' = x + y d\theta_z - z d\theta_y \;, \tag{3a}$$
$$y' = y + z d\theta_x - x d\theta_z \;, \tag{3b}$$
$$z' = z + x d\theta_y - y d\theta_x \;, \tag{3c}$$
$$t' = t. \tag{3d}$$

c) Scale

A variable x is rescaled by multiplying it by a scale factor. In general each coordinate can be scaled independently. Here, I will take for simplicity the same scale change for all the variables. An infinitesimal scale factor has the form 1 + dk, yielding

$$x' = x + dkx \tag{4a}$$
$$y' = y + dky \tag{4b}$$
$$z' = z + dkz \tag{4c}$$
$$t' = t + dkt \tag{4d}$$

d) Velocity

A small change in velocity remaps the spacetime variables. The Galilean transformation does not affect time. The more general Lorentz transformation remaps both the spatial and temporal variables. For a very small velocity differential the Lorentz transformation gives

$$\vec{r}' = \vec{r} - (\vec{dV})t \;, \quad t' = t - \frac{1}{c^2}\left(\vec{dV} \cdot \vec{r}\right)$$

JOHNSON

and so the velocity transformations are

$$x' = x - dV_x t ,$$ (5a)

$$y' = y - dV_y t ,$$ (5b)

$$z' = z - dV_z t ,$$ (5c)

$$t' = t - \frac{1}{c^2}\left(dV_x x + dV_y y + dV_z z\right).$$ (5d)

The total transformation is obtained by combining the infintesimal changes:

$$x' = x + \epsilon_x$$

$$y' = y + \epsilon_y$$

$$z' = z + \epsilon_z$$

$$t' = t + \epsilon_t$$

where

$$\epsilon_x = dx + dkx - dV_x t + yd\theta_z - zd\theta_y$$

$$\epsilon_y = dy + dky - dV_y t + zd\theta_x - xd\theta_z$$

$$\epsilon_z = dz + dkz - dV_z t + xd\theta_y - yd\theta_x$$

$$\epsilon_t = dt + dkt - \frac{1}{c^2}\left(xdV_x + ydV_y + zdV_z\right)$$

This shows that the general infinitesimal affine coordinate transformation

$$\epsilon_\nu = dx_\nu + \sum_\mu d\alpha_{\mu\nu} x_\mu$$ (6)

can be interpreted as a joint transformation under translation, rotation, scale, and velocity changes. A first order Taylor's series expansion of f' is

226

$$f' = f + \sum_{\mu} \epsilon_{\mu} \frac{\partial f}{\partial x_{\mu}}. \tag{7}$$

When the expressions for ϵ_{μ} are used, and the terms regrouped, then in vector notation

$$f'(x', y', z', t') = f(x, y, z, t) + dt\frac{\partial f}{\partial t} + d\vec{r} \cdot \vec{\nabla}f + d\vec{\theta} \cdot (\vec{r} \times \vec{\nabla}f)$$

$$+ dk \left(t\frac{\partial f}{\partial t} + \vec{r} \cdot \vec{\nabla}f \right) - d\vec{V} \cdot \left(\frac{\vec{r}}{c^2}\frac{\partial f}{\partial t} + t\vec{\nabla}f \right). \tag{8}$$

This is the basic result of this paper. It shows that the change in the function is composed of four transformations, each involving local gradient operations and the original coordinate variables acting on the original untransformed function.

ATR

This result will now be applied to the problem of remapping the perceived scene for an automatic target recognition system. It will serve as an image preprocessing algorithm for a matched filter correlator. Identify f and f' as two successive frames from a missile-mounted video camera whose optical axis is parallel to the velocity vector of the missile (Figure 1). Let Δz be the distance travelled by the missile in frame time τ and let f_0 be the focal length of the camera lens. Assume for simplicity there is no acceleration: $d\vec{V} = 0$. Since a single frame from the camera is a fixed image, $\frac{\partial f}{\partial t} = 0$ as well. The change in the image scale is dk = $\Delta z/z$ where z is the on-axis distance to the plane containing the object at each pixel. Assume that the missile rotation is given by
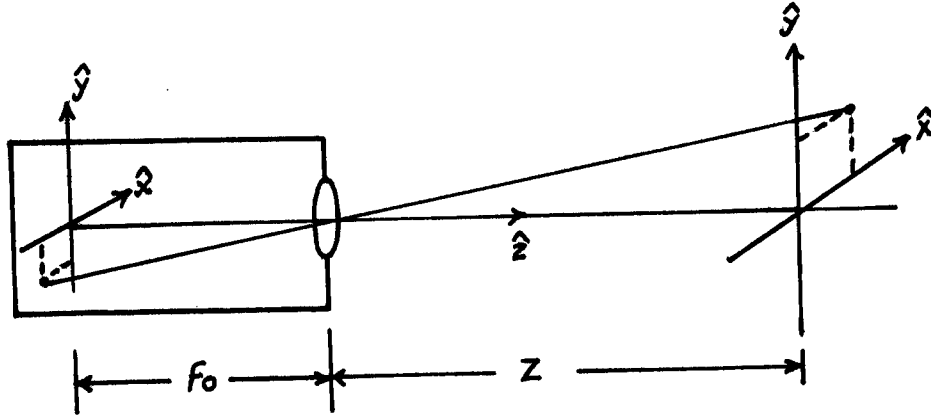
Figure 1. Geometry for a missile-mounted camera.

$\vec{\omega}_M$ and that J rotating objects in the scene have their rotation centers offset by $\vec{R}_J$. Scene objects can be rotating and/or translating, and their transverse translations at range z are $\vec{V}_K\tau$ and the object sets K and J may overlap. Set the origin at the center of expansion $\vec{r}_0 = 0$. Note that the object-image transverse coordinates are reversed and also scaled by the ratio of the object and image distances. For $z \gg f_0$ this ratio is essentially f_0/z. Eq. (8) then becomes

$$f' = f - \frac{f_0\tau}{z}\sum_K \vec{V}_K \cdot \vec{\nabla}f + \tau\sum_J \vec{\omega}_J \cdot \left[\left(\vec{r} + \frac{f_0}{z}\vec{R}_J\right) \times \vec{\nabla}f\right]$$

$$+ \frac{\Delta z}{z}\vec{r} \cdot \vec{\nabla}f + \tau\vec{\omega}_M \cdot (\vec{r} \times \vec{\nabla}f)$$

$$(9)$$

where $\vec{r}$, $\vec{\nabla}$ are in the image frame coordinates and z, $\bar{R}_J$, and $\vec{V}_T$ are in the object frame coordinates. The parameters Δz and $\vec{\omega}_M$ are

constant for every pixel. The parameters z, $\vec{\omega}_J$, $\vec{R}_J$, and $\vec{V}_T$ may have different values at every pixel. For opaque objects, at a given pixel there can be only one object. If $\vec{\omega}_M$ is confined to roll only, then it will produce a global rotation effect which is the same at every pixel. As will be shown in the next section, missile pitch and yaw rates produce apparent translations of the entire scene. Again, their effect is global and is the same at every pixel. Thus missile rotations can be distinguished from other translations and rotations by their global effect, an intuitively obvious result made quantitative by Eq. (9).

If the objects are rotating about axes normal to the z-axis, the perceived effect is that of a velocity gradient across the viewed area of the object. This is equivalent to an acceleration and requires the use of the Lorentz factor. Since it was assumed to be zero, we are restricting this case to object rotations about the z-axis. For a single object undergoing transverse translation and rotation as viewed by an incoming nonrotating missile, the frame difference is

$$T = f' - f = \frac{-f_0 \tau}{z} \left(V_x - \omega R_Y \right) \frac{\partial f}{\partial x}$$

$$- \frac{f_0 \tau}{z} \left(V_y + \omega R_x \right) \frac{\partial f}{\partial y}$$

$$+ \omega \, \tau \left(x \frac{\partial f}{\partial y} - y \frac{\partial f}{\partial x} \right)$$

$$+ \frac{\dot{\Delta z}}{z} \left(x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} \right) \tag{10}$$

The frame difference T and the four derivative functions are all computed locally from the two frames f and f'. Eq. (10) thus has the

general form of

$$T(x,y) = \sum_{j=1}^{4} \alpha_j(x,y) \, I_j(x,y)$$

(11)

where $I_1 = \dfrac{\partial f}{\partial x}$, $I_2 = \dfrac{\partial f}{\partial y}$, $I_3 = x\dfrac{\partial f}{\partial y} - y\dfrac{\partial f}{\partial x}$, $I_4 = x\dfrac{\partial f}{\partial x} + y\dfrac{\partial f}{\partial y}$.

and the α_j are two dimensional maps of the object features of $\vec{V}_T$, $\vec{\omega}$, $\vec{R}_c$, and z. The coefficient α_4 is a map of the inverse z-axis distance to each object plane. The rotation map is α_3. The coefficients α_1 and α_2 involve the target velocity and/or center of rotation. The α_j are a good choice of object features, for they are constant over the area of each object (opaque rigid body), they do not depend on intensity, and they give quantitative object information as well as locational and grouping decision data.

Feature Extraction

Consideration of training algorithms using adaptive relaxation yields the following comments: we have a training signal composed of a linear superposition of four input signals for every pixel. We wish to obtain the superposition coefficients of $\alpha_j(x,y)$. If we already knew the answer, we could use a simple linear neural network to adapt the input weights $m_j(x,y)$. But all we have is the weighted sum $T(x,y)$ and the four input signals $I_j(x,y)$. On the other hand, we argue that for a rigid body of finite spacial extent the corresponding features will be constant in the area subtended by its image. This suggests the requirement for a training algorithm in which the weight adaptation rule involves not only the value of the weight and/or input at a given pixel but also their values in a local neighborhood of that pixel.

Consider a linear minimization algorithm in which an estimate S of the frame difference T is formed, based on estimates m of the weights α. For simplicity let the single index n indicate the nth pixel. Eq (11) is then

$$T_n = \sum_j \alpha_{nj} I_{nj}$$

(12)

and its estimate is

$$S_n = \sum_j m_{nj} I_{nj}$$

(13)

The error is $\epsilon_n = T_n - S_n$ and its energy function is $\sum_n \epsilon_n^2$. Add the smoothness criteria in the form of the local variance of the weight estimates to obtain the total energy function

$$E = a\sum_n \epsilon_n^2 + b\sum_n \sum_j \sigma_{nj}^2$$

(14)

where

$$\sum_n \sum_j \sigma_{nj}^2 = \left(\frac{1}{2K}\right)\sum_n \sum_j \sum_{k=n-K}^{n+K} \left(m_{kj} - \overline{m}_{nj}\right)^2$$

(15)

and

$$\overline{m}_{nj} = \left(\frac{1}{2K+1}\right)\sum_{l=n-K}^{n+K} m_{lj} \cdot$$

(16)

and a, b are constants.

We will obtain a weight iteration equation by the usual requirement[7] of

231

$$\frac{\partial E}{\partial m_{pq}} = -m_{pq} .$$

(17)

The frame difference error term yields the expected term of

$$-2a(T_p - S_p) I_{pq}.$$

(18)

Application of Eq (17) to Eqs (15) and (16) yields

$$\frac{1}{2K}\sum_n\sum_j\sum_k 2\left(m_{kj} - \overline{m}_{nj}\right)\left(\delta_{pk} - \left(\frac{1}{2K+1}\right)\sum_l \delta_{pl}\right)\delta_{qj}$$

(19)

where δ_{pk}, δ_{pl}, and δ_{qj} are Kronecker deltas. This breaks down into four terms. The δ_{qj} factor pulls out the term $j = q$ from the j-sum. When the terms are evaluated we obtain the weight update equation

$$\dot{m}_{pq} = 2a (T_p - S_p) I_{pq} + [(2K + 1)/K] b (\overline{\overline{m}}_{pq} - m_{pq}).$$

(20)

where $\overline{\overline{m}}_{pq}$ is the double average of m_{pq}.

The first term selects those weights which can satisfy Eq (12) while the second term restricts them to the subset of weights which are locally smooth from one pixel to the next. The double average term samples the weights in a neighborhood of size $4K + 1$ and linearly weights their contribution in decreasing proportion to their distance from the center. For example, if $K = 1$ then two applications of Eq (16) give (the q-subscript is suppressed for clarity)

$$\overline{\overline{m}}_p = \frac{1}{9} (m_{p-2} + 2m_{p-1} + 3m_p + 2m_{p+1} + m_{p+2})$$

Eqs (12), (13), and (20) comprise the actual algorithm. The weight update equation (Eq (20)) began with a simple LMS error form[8] and received a relaxation term and an additive local neighborhood excitation term. The relaxation term gives it the appearance of the neural Hebbian associative memory learning law[9], but the double average term has a distinctly nonbiological flavor since it requires the transport of weight values from one node to another. Backpropagation nets were also categorized as nonbiological due to this same weight transport requirement, but recently it has been shown[10] possible to formulate networks which conceptually eliminate this objection. Basically, these nets contain signals proportional to the weights. Since signals can be propagated, neural architectures can be devised.

The same approach can be used here. Eq (13) shows a sum of products. Mathematically it does not matter if we use the I_{nj} as weights and the m_{nj} as signals rather than the other way around. Figure 2 shows the neural net architecture for $j = 1,2$. The nodes of the estimate slab have connecting weights of strength I_{nj}, receive input signals m_{nj}, and sum over j to form the linear output estimate S_n. The error slab differences T_n and S_n to form the error distribution ϵ_n. It is replicated and sent to each of the j update slabs where it is weighted by I_{nj} and summed with the double average $\overline{\overline{m}}_{nj}$. The outputs go to the averaging slabs, and also go to the estimate slab to complete the cycle.

Since the entire system is linear (no sigmoid functions) it is an attractive candidate for implementation in SLM optics or in electronic pipelined hardware.

Discussion
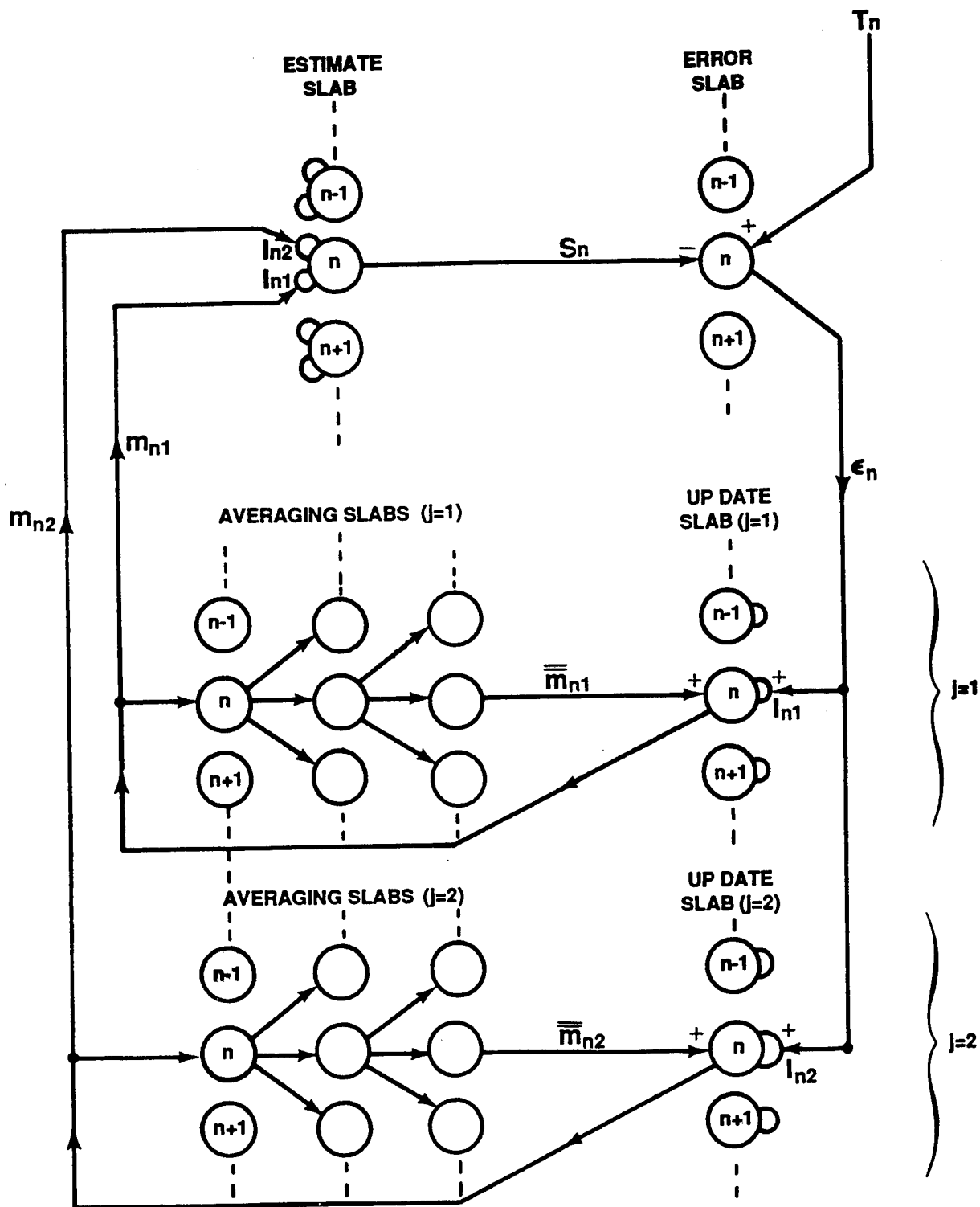
Figure 2.   Linear Neural Net for the Adaptive Weights.

It has been shown that an infintesimal contact transformation under translation, rotation, scale and velocity changes is sufficient to specify a complete infinitesimal affine transform of the spacetime coordinates. It leads to an expression giving the transform of a function of spacetime in terms of the untransformed coordinates and partial derivatives of the original function. When applied to dynamic imagery, this expression gives an explicit statement of an image in terms of object features where the features are maps of object velocity, range, rotation and acceleration. Algorithms for computing these feature maps are discussed. The requirements for a relaxation training algorithm are shown to include a need for adaptive terms containing not only the information at a single point or processing node but also information from (at least) a local neighborhood. Such an algorithm is derived.

The transformation for dynamic imagery from a missile-mounted camera can distinguish between missile rotations and object rotations. Missile roll produces a global rotation. The resulting rotation map α_3 then has a constant offset added to every pixel. Object rotations show up as local distributions in the rotation map. The missile roll appears as an additive bias term. The other two missile rotations will appear as global translations in the maps of α_1 and α_2 containing translation information. This can be seen by inspecting the above result for a rotating surveillance camera. Again, these global factors are a bias term added to the object distributions. Object rotations around an axis normal to the optical axis of the viewing camera will produce localized optical flow distributions bounded by the area of the object itself.

The forward velocity of the missile produces a continually rescaled image. The transformation equation contains a map α_4 of the scale change at every pixel. This is map of the inverse normal range to the object imaged at that pixel. The information in this map provides three-dimensional data which could be used to generate SAR-like views of the scene. These are desired invariant imagery

appropriate for input to a matched filter correlator. Alternately, the inverse range map information could be used to synthetically rescale local regions of the actual image to a standard range. It is also noted that a binocular viewing system provides a much simpler range algorithm.

References

1. Crowe, W. M. and Kirsch, J. C., "Optical Correlator Guidance Technology Demonstration," Proc. SPIE Vol. 938 Digital and Optical Shape Representation and Pattern Recognition, p 29-35, 4-6 April 1988.

2. Duthie, J. G. and Upatnieks, J., "Compact Real-Time Coherent Optical Correlators," Opt. Eng. Vol. 23, No. 1, Jan/Feb 1984, p 007-011.

3. Gregory, D. A. and Liu, H. K., "Large-Memory Real-Time Multichannel Multiplexed Pattern Recognition," App. Opt. Vol. 23, No. 24, 15 Dec 1984, p 4560-4570.

4. Verri, A., Girosi, F., and Torre, V., "Mathematical Properties of the Two-Dimensional Motion Field; from Singular Points to Motion Parameters," J. Opt. Soc. Am. A, Vol. 6, No. 5, May 1989, p 698-712.

5. Goldstein. H., Classical Mechanics, Addison-Wesley, Reading, MA, Chap. 8, Section 8-6 (1950).

6. Leighton, R. B., Principles of Modern Physics, McGraw-Hill, New York, NY, Chap. 1, Section 1-11 (1959).

7. Rumelhart, D. E. and McClelland, J. C., Parallel Distributed Processing, MIT Press, Cambridge, MA, Vol. I & II (1986).

8. Widrow, B. and Hoff, M. E., "Adaptive Switching Circuits," IRE WESCON Conv. Rec., Part 4, 96 (1980).

9. Grossberg, S., "Nonlinear Neural Networks: Principles, Mechanisms, and Architectures," Neural Networks, Vol. 1, No. 1, p 17-62 (1988).

10. Hecht-Nielsen, R., "Theory of the Backpropagation Neural Network," INNS/IEEE Proc. 1JCNN, Vol. I, p 593-606 (1989).

Xenon Chloride Laser Scaling

Thomas H. Johnson, COL, EN
*Harry E. Cartland, CPT, CM
Science Research Laboratory
United States Military Academy
West Point, New York 10996-5000

Introduction

The past decade has seen a rapid increase in interest in the scaling characteristics of various types of laser systems. This interest was spurred to a large extent by the announcement of the Strategic Defense Initiative in 1983. Under this program, directed energy weapons of very high power were envisioned to play a key role in a non-nuclear defense against a foreign ballistic missile threat. Much of the early optimism towards attainment of this goal has been blunted by the sheer magnitude of the technical obstacles encountered. In fact many of these difficulties are not simply engineering problems; they involve basic issues in the physical sciences which have yet to be properly addressed, if at all.[1] Still, considerable progress has been made which suggests that some perhaps less ambitious goals may be realized in the next decade.

One area in which attention has been focussed is the development of rare gas halide excimer lasers. Because of their combination of wavelength (308nm) and efficiency (>4%), xenon chloride (XeCl) lasers have been examined as a potential cornerstone of a ground based laser (GBL) system for ballistic missile defense. Though XeCl lasers, and competing technologies for that matter, have not lived up to early expectations in this regard, other more realistic applications for these devices seem possible in the near-term. Construction of a XeCl laser with sufficient power to fulfill a low earth orbit ASAT (antisatellite) mission is quite possibly within the realm of current engineering technology. Considerably less power is necessary for a number of promising optical imaging applications that rely on a source of coherent light. At a more basic level, a high power XeCl test laser would permit a number of important atmospheric propagation issues to be studied, issues that are common to all candidates for the GBL role.

To this point our concern has been exclusively with military applications, but there are others as well. For instance in medicine excimer lasers, more so than their longer wavelength counterparts, have shown promise as surgical tools.[2] In this case the emphasis is on scaling XeCl lasers to longer pulse lengths rather than higher power. Excimer lasers may also see some use in the fabrication of microelectronics components,[3] and remain unsurpassed in many laboratory applications requiring an intense source of UV photons. On a larger scale, excimer lasers are serious contenders as drivers for inertial confinement fusion.[4]

Efficient development of XeCl lasers to fill these various roles presupposes a knowledge of the fundamental chemical and physical processes governing their performance. An understanding of the basic underlying principals provides a foundation from which the operational properties of new devices may be predicted. Over the past five years we have developed a comprehensive kinetic and extraction model of XeCl laser performance which incorporates these principals and satisfactorily explains the wide range of extant data. Here we use our model, the details of which can be found elsewhere,[5] to examine the scaling characteristics of electron beam (e-beam) excited XeCl lasers.

We begin by presenting a brief description of our model which includes an examination of a number of basic kinetics issues. Consideration is also given to validation of the code by comparison of model predictions to experimental data. We then turn to applications of the model to examine the scaling characteristics of XeCl lasers with gas mixture, pump power, gain length, and pulse length.

The Model

The current model comprises 41 distinct species which interact through 202 chemical and physical processes. For HCl, the chlorine donor, only the ground and two vibrationally excited states are considered explicitly since further collisional excitation into $v > 2$ cannot compete effectively with electron dissociative attachment in the vibrationally "hot" molecules.[6] The Xe manifold consists of the ground state, the ion, and three intermediate electronically excited species which represent groups of atomic levels clustered roughly at the energies of the 6s, 6p, and 4f atomic states. Simulations involving both Ar and Ne buffer gases are possible, and two metastable electronic and one ionic state of these species are considered. For the high pressure conditions under which e-beam pumped XeCl lasers are run, typically with 2 atm of Ar or 4 atm of Ne buffer gas, it is necessary to include in the model processes involving homonuclear and heteronuclear diatomic rare gas metastables and ions. In the XeCl system lasing occurs between the $^2\Sigma$ B and X states, but the accuracy of the simulation depends on inclusion of the $^2\Pi$ C (bound) and A (repulsive) states with which they are collisionally mixed. Secondary electrons play a crucial role in plasma kinetics, and rate constants for most electron processes are calculated from temperature models for which a Boltzmann distribution of kinetic energies (typically 2 eV) is assumed.[7] The time dependent number densities of the species mentioned above, and a number of others as well, are determined by a FORTRAN integrator which simultaneously solves the set of coupled rate equations maintaining detailed balance in the model.

An arbitrary e-beam current pulse profile serves as the source of input energy through ionization and electronic excitation of rare gas atoms. Cross sections for these processes at the primary (beam) electron kinetic energy of interest are interpolated from compilations found in the literature.[8] Energy extraction from the medium is accomplished with a Rigrod Model which is

suitable for conditions of high gain and homogeneous broadening.[9]

Several rate constants and cross sections which have a major impact on the model's performance have been the subject of much controversy in past years. Our code uses a stimulated emission cross section of 4.2×10^{-16} cm^2 and an absorption cross section of 1.2×10^{-16} cm^2; these values were calculated from a detailed partition function[5] using spectroscopically determined parameters[10] and verified by fits to actual XeCl B → X gain measurements[11]. Vibrational excitation of HCl through secondary electron collisions significantly enhances the important HCl-electron dissociative attachment process feeding the ion channels that lead to formation of the upper lasing level; simulations using a $\Delta v = 1$ rate constant of 1.5×10^{-9} cm^3 s^{-1}, which incorporates both electron excitation and superelastic de-excitation effects, accurately track experimental measurements of HCl excited state number densities[12]. A Stern-Volmer analysis of XeCl(B) fluorescence data[13,14], which the code accurately simulates, shows that electron quenching of the upper level is important only for very low HCl concentrations (<1 Torr); we have adopted a value of 2.0×10^{-8} cm^3 s^{-1} for this rate constant in agreement with previous analyses[15,16]. Finally, free-bound photoassociation (Eqn.(1)) and reconstitution of the ground state (Eqn.(2))

$$Xe + Cl + h\nu \rightarrow XeCl(B) \qquad (1)$$

$$Xe + Cl + M \rightarrow XeCl(X) + M \qquad (2)$$

play important roles in limiting extraction efficiency; for the former process a value of 1.0×10^{-37} cm^2 agrees with estimates found elsewhere[17], while a three body rate constant of 1.2×10^{-33} cm^6 s^{-1} for the latter was determined from the equilibrium constant[18] and the rate constant for the reverse (ground state quenching) process[19]. All code validation and scaling simulations use the _same_ set of rate constants.

The model was validated by simulation of a wide range of experimental data. Unlike previous models of the XeCl laser system, this code accurately reproduces the observed HCl, Xe, buffer gas, and power scaling of e-beam pumped devices operating in both the "hard" pump (~3 MW/cm³) short pulse (~80 ns) regime represented by the Sandia experiments[20], and the "soft" pump (~100 kW/cm³) long pulse (2.5 μs) regime represented by the AERL experiments[21]. The Sandia simulations are shown in Figure 1. In intermediate operating regimes, represented by experiments conducted at Thermo Electron Technologies[22], Northrop[23], NRL[24], and Spectra Technologies[25], data is available on the time dependent behavior of various kinetic species such as electrons, vibrationally excited HCl, XeCl(B), and Xe_2Cl^*. These results are a more important test of the model's validity than the gross scaling features mentioned earlier, and the code simulates them faithfully as well.
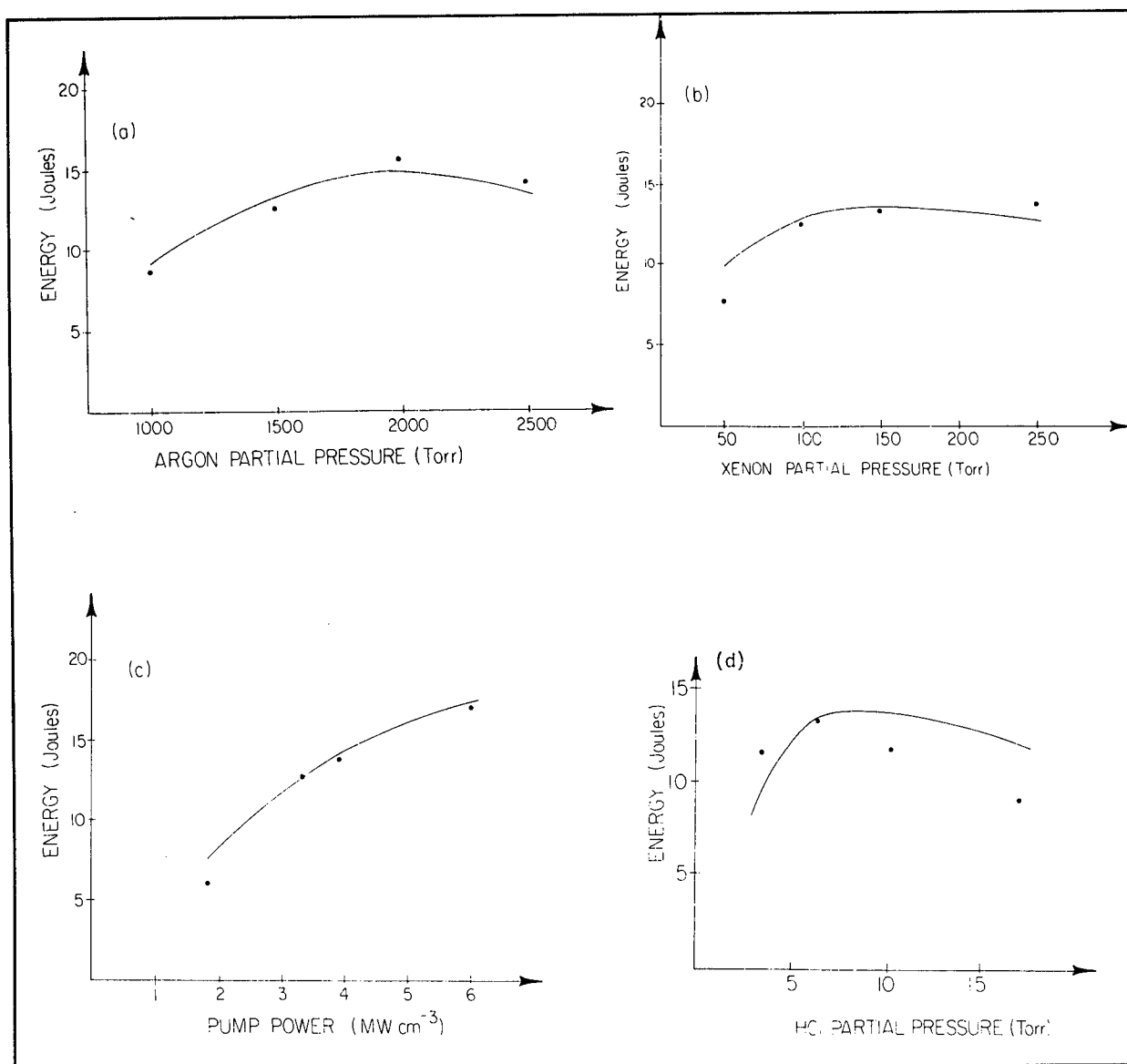
Figure 1. Data (points) and model simulations (solid lines) for scaling of energy out (joules) in the Sandia experiments (Ref. 20) vs (a) argon pressure (Torr), (b) xenon pressure (Torr), (c) pump power (MW/cm³), and (d) HCl pressure (Torr).

Scaling Calculations

Before scaling calculations are considered, it is instructive to review the basic chemical and physical processes leading to formation of the upper

lasing level. Understanding them will aid in the interpretation of the computer simulations which follow.

Regardless of pumping regime, about 75% of the B state population is derived through a series of reactions which constitute the "ion channel". As shown below in Eqn.(3), Cl^- is produced by dissociative attachment to HCl, a process enhanced on vibrational excitation by several orders of magnitude.

$$HCl(v = 0,1,2) + e^- \to H + Cl^- \tag{3}$$
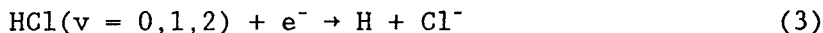
Xenon atomic ions and xenon containing molecular ions, both of which are important, are produced principally through charge transfer reactions involving buffer gas ions generated by the e-beam (Eqn.s(4)).

$$Ar^+ + 2Ar \to Ar_2^+ + Ar$$
$$Ar_2^+ + Xe + M \to Xe^+ + 2Ar + M$$
$$Xe^+ + Xe + M \to Xe_2^+ + M \tag{4}$$
$$Ar^+ + Xe + M \to ArXe^+ + M$$
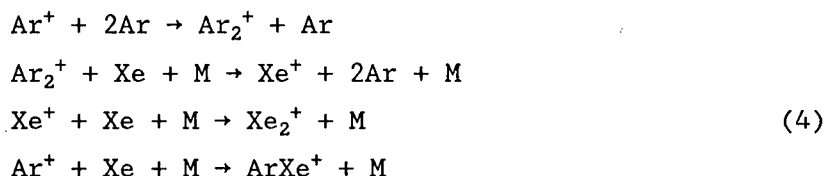
Here M is any third body. Neutralization reactions then feed the upper laser level as Eqn.(5) shows.

$$Xe^+, Xe_2^+, ArXe^+ + Cl^- \to XeCl(B) + \ldots \tag{5}$$

The process is shown for an Ar buffered system, but analogous reactions occur with Ne.

In addition to the ion channel, the "neutral channel" also populates the B state largely through the harpooning reactions in Eqn.(6).

$$Xe^{**}, Xe^{***} + HCl(v = 0,1,2) \to XeCl(B) + H \tag{6}$$
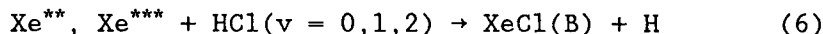
Displacement reactions are also possible in Ar buffered systems, Eqn.(7),

$$Xe + ArCl^* \to XeCl(B) + Ar \tag{7}$$
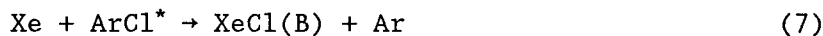
but not with Ne buffer gas since $NeCl^*$ predissociates. It should be emphasized at this point that the preceding discussion presents a rather simplified picture of a very complicated problem.

The calculated curves in Figure 2 show the power scaling of small signal gain (g_o) in Ar and Ne buffer gases under typical lasing conditions.
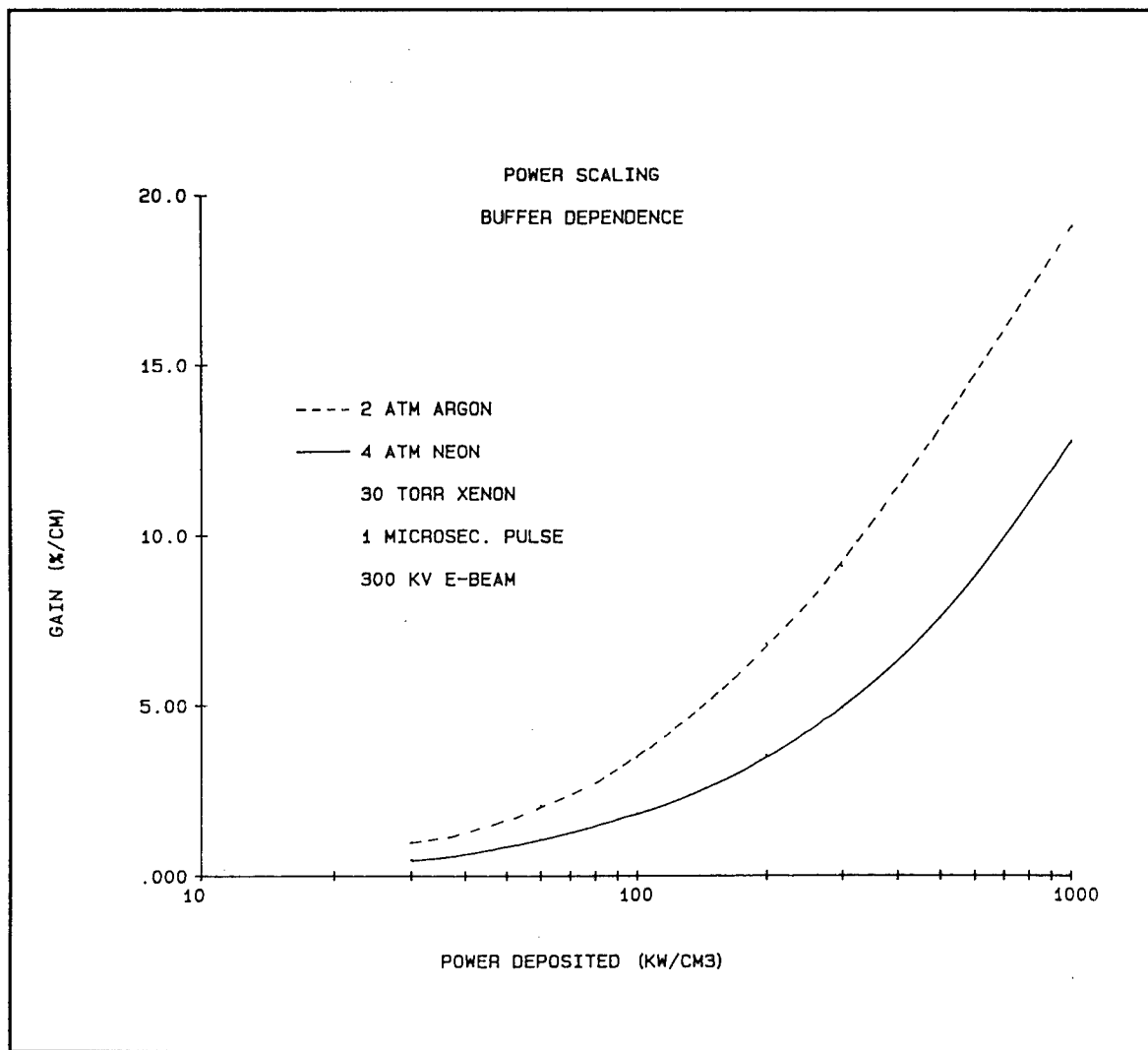
Figure 2. Calculated small signal gain (%/cm) for argon and neon buffer gases under typical lasing conditions. HCl concentration was optimized at the deposition energy.

Note that the NRL gain measurements[24] which the code predicts within experimental error cover much of the same range in power. Based on Figure 2 one is tempted to conclude that laser performance will be uniformly better with an Ar buffer, and possibly to attribute it to reaction pathways available in Ar but not Ne buffered systems. This figure also suggests that the potential to extract energy from an e-beam excited XeCl laser is only limited by the technology required to pump it. In fact, none of these statements, especially the last one, are correct.

The existing experimental evidence indicates that extraction efficiencies are slightly better in Ne than in Ar. And a closer examination of the kinetics reveals that the displacement reaction, Eqn.(7), is very slow; under typical conditions the characteristic reaction time of ArCl* with Xe is almost a factor of 8 longer than its spontaneous radiative lifetime. Finally our pump power scaling calculations and the experimental data attest to the falsity of the last conclusion. (Compare, for example, the curvatures of Figure 1c and Figure 2.) Clearly unloaded gain in itself does not constitute a valid predictor of laser performance.
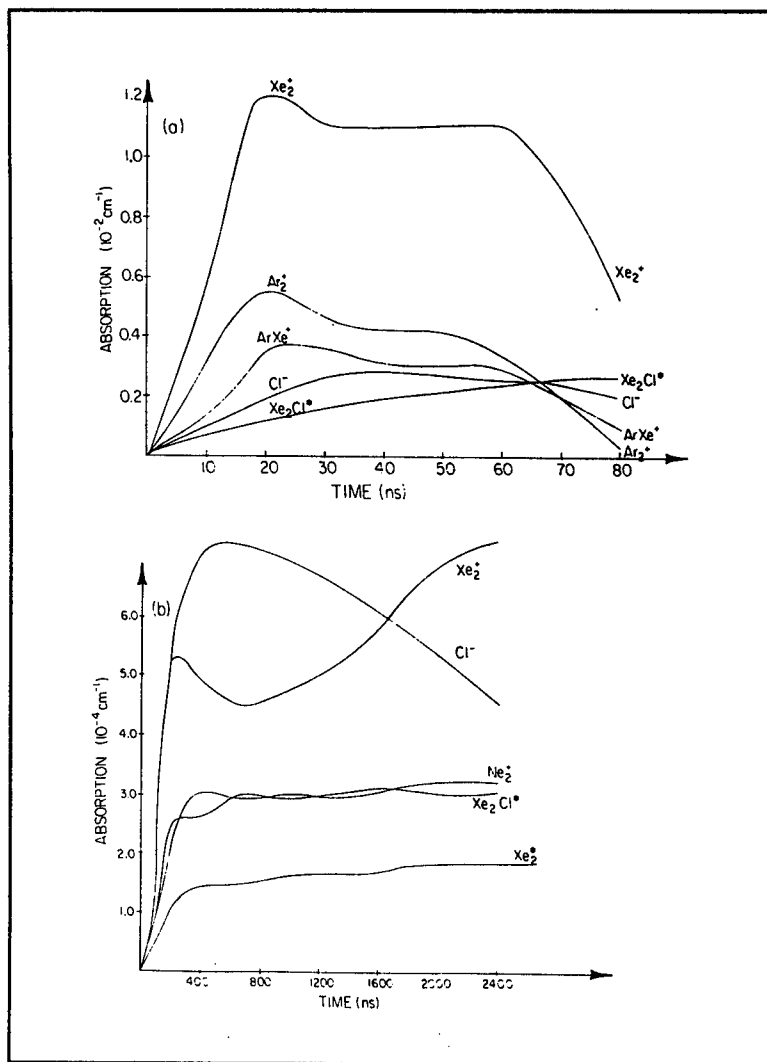
Figure 3. Model simulations of transient absorptions (cm^{-1}) caused by the principal absorbing species vs laser pulse time (ns). (a) the "hard" pump, short pulse Sandia experiments (Ref. 20). (b) the "soft" pump, long pulse AERL experiments (Ref. 21).

What has been neglected so far in this analysis is the effect of absorption. As shown in Figure 3, absorption is dominated by atomic and

molecular ions throughout the pulse under both "hard" and "soft" pumping conditions. Hence the very same species which contribute most to population of upper laser level are also responsible for most of the absorption losses. The mutually ionic nature of pumping and absorption is the key to understanding XeCl scaling.

A simple expression relating gain, absorption, and extraction can be derived. The differential increase in intracavity intensity (I) depends on the difference between the gain (g) and absorption (α) coefficients as shown below in Eqn.(8).

$$dI = (g - \alpha)Idz \tag{8}$$

Maximum intensity is obtained for dI/dz = 0. Solving for I and using the usual expressions for the gain coefficient and saturation intensity in a homogeneously broadened system, Eqn.s(9),

$$g = \frac{g_0}{1 + I/I_{sat}}$$

$$\tag{9}$$

$$I_{sat} = \frac{h\nu}{\sigma_s \tau}$$

one finds that the power extracted from the medium is a function of the small signal gain to absorption ratio (Eqn.(10)).

$$I_{ext} < I = (\frac{g_0}{\alpha} - 1)I_{sat} \tag{10}$$

Here σ_s is the stimulated emission cross section, ν the transition frequency, τ the upper state lifetime, and I_{ext} the extracted intensity which is less than the intracavity intensity by a factor of the fraction output coupled (FOC).

Figure 4 shows the model's predicted power scaling for a typical electron beam pumped device with both Ar and Ne buffer gases. The abscissa, g_0/α, is now a much improved indicator of laser performance. Neon is expected to out-perform Ar as a buffer gas, in agreement with experimental observations. The calculations also reveal an optimum deposition power, in this case ~400 kW/cm³, for the conditions modelled. The maximum in g_0/α is an indirect result of the complex interplay between the processes leading to population of the upper laser level and those leading to absorption losses. At high deposition rates increased gain is overcome by high absorber number densities. In addition, the model predicts a weak dependence of g_0/α on Xe partial pressure (not shown) over the typical range of several tens of Torr, but the general features of the power scaling are not significantly altered.
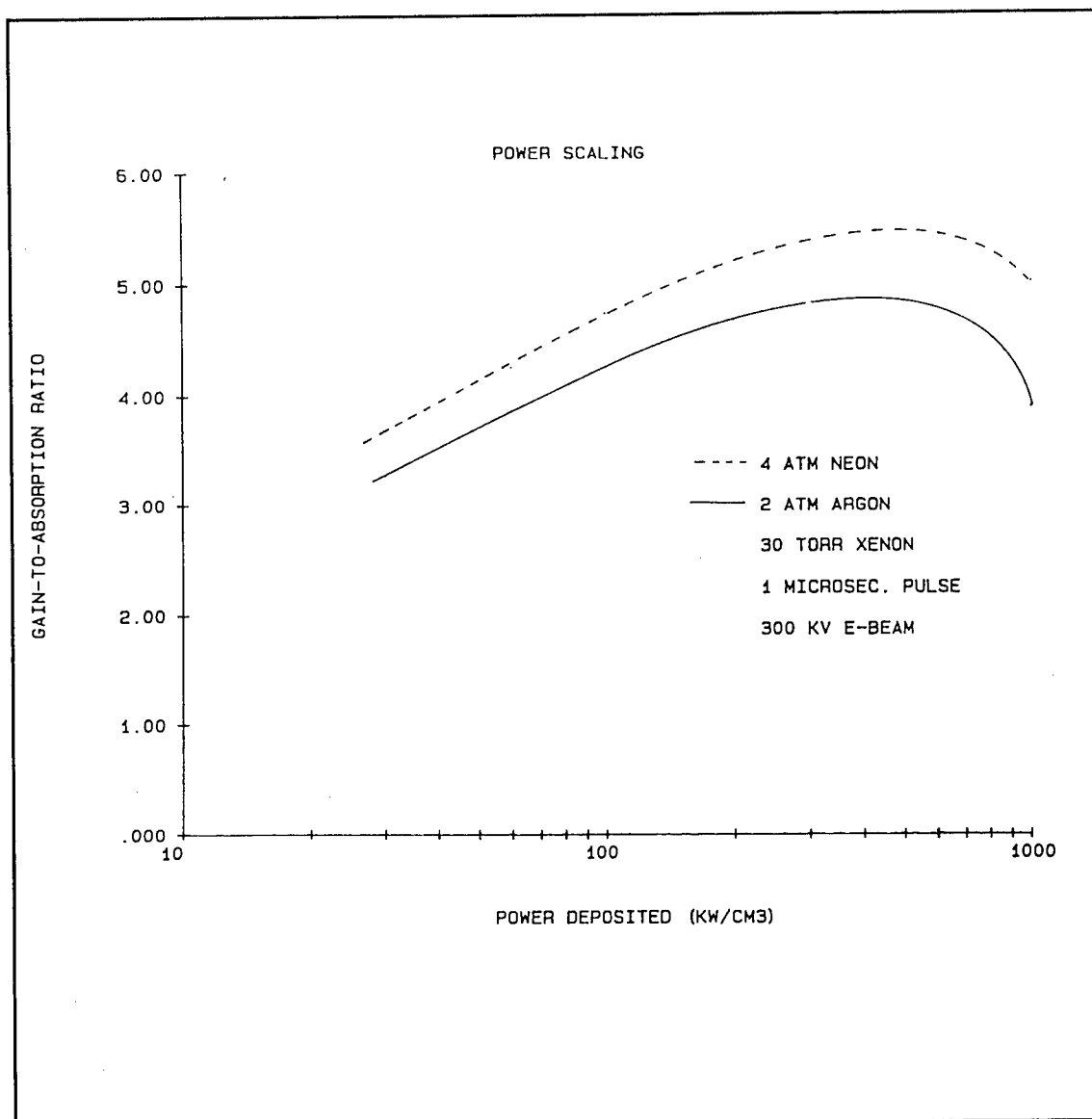
Figure 4. Calculated small signal gain to absorption ratio for argon and neon buffer gases. Conditions are the same as those in Figure 2. HCl concentration was optimized at the deposition energy.

Xenon chloride lasers operating in the "soft" pump, long pulse regime have been of particular interest because of the more conservative pulse power technologies required to pump them. Although peak powers cannot match those obtainable with hard pumping, the lower gain in this case reduces the detrimental effects of amplified spontaneous emission thus making larger

apertures and longer gain lengths practicable. A properly constructed device operating in this regime will then be able to equal or exceed in total energy per pulse devices operating under "hard" pump short pulse conditions.

In Figure 5a the code calculates the integrated energy fluence from a hypothetical 1m "soft" pump (80 kW/cm^3) laser as a function of pulse length. The HCl partial pressure was carefully optimized for each pulse length to prevent halogen donor burn up prior to pump termination while minimizing the effects of HCl quenching in the upper lasing level. Note that local efficiencies (energy extracted/energy deposited) of nearly 5% are possible for short pulse lengths. Figure 5b shows the gain length scaling of the same laser except that the pulse length was limited to 2 μs since, as shown in Figure 5a, little energy is gained by extending it further. At 15m almost 4 J/cm^2 is extracted with potential remaining for further improvement at even longer gain lengths. For a typical 1m^2 aperture, this corresponds to a pulse energy of ~40 kJ, or a peak power of ~20 GW. Of course even higher pulse energies could be obtained through the use of a series of single pass amplifier modules.

Similar pulse and gain length scaling calculations (not shown) were performed for a laser designed to operate in the "moderate" pump regime (~300 kW/cm^3) under the following conditions: 675 keV e-beam energy, 2 atm Ar, 45 torr Xe, and optimized HCl concentration and FOC. In this case peak local efficiency is near 4% for a 1m gain length. The predicted energy fluence is also down, by 25%, to ~3 J/cm^2 at 15m, and shows a significantly decreased enhancement for gain lengths longer than 5m as compared to the previous case.

Conclusion

In order for excimer lasers to realize their full potential in a variety of military, industrial, medical, and scientific applications, the fundamental physical processes determining their performance properties must be fully understood. To this end, we have developed a comprehensive kinetic and extraction model for electron beam excited XeCl lasers that accurately reproduces experimental data over a wide range of operating conditions. The model has allowed us to demonstrate that scaling of these devices depends on a complex interaction of ionic processes which determine the rates of upper state pumping and absorption loss. Hence the quantity g_o/α must be used to predict laser scaling characteristics. Finally we have shown that pulse energies of tens of kilojoules, or peak powers of tens of gigawatts, are obtainable with current, conservative pulse power technology.
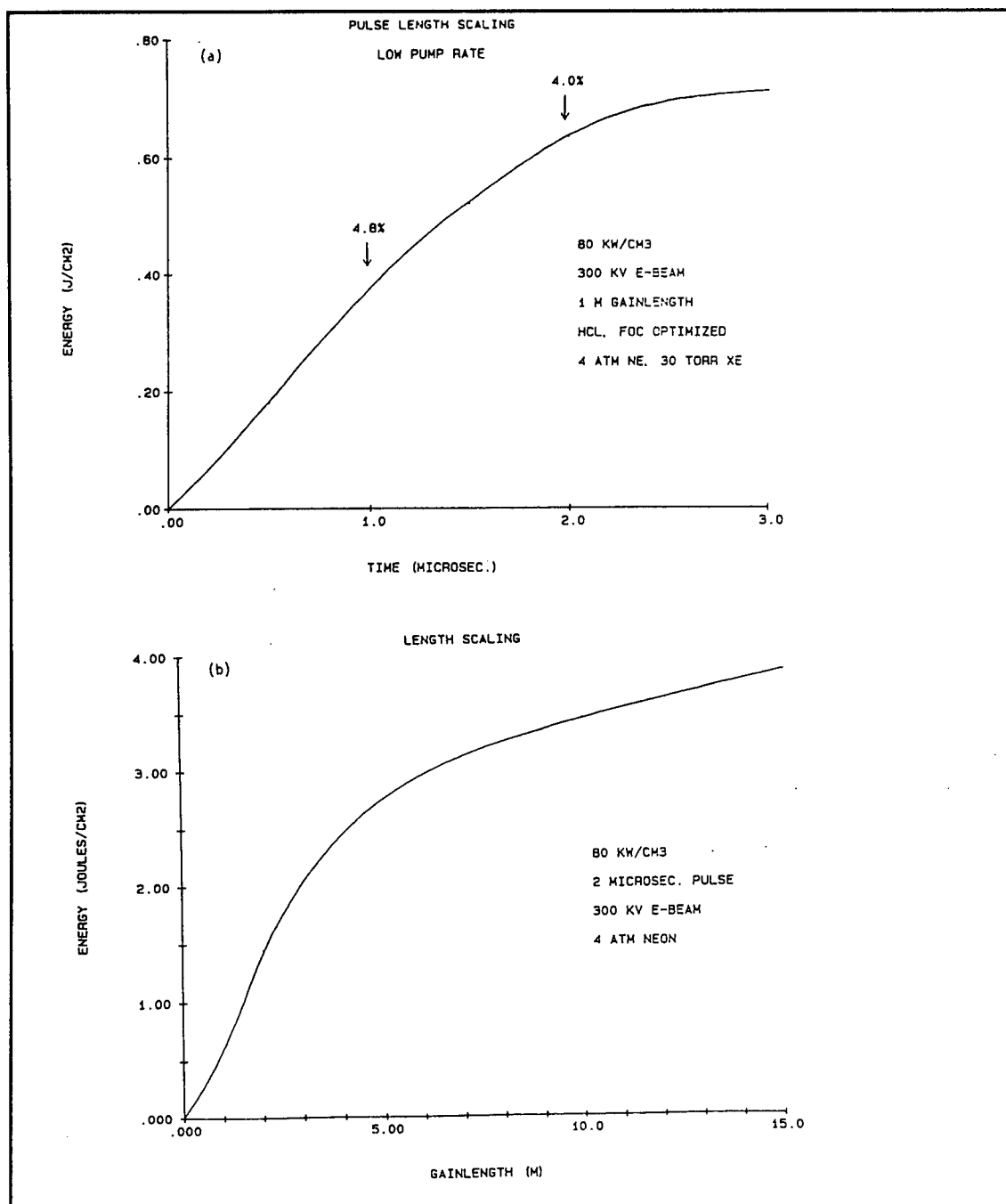
Figure 5. Calculated energy fluence (J/cm^2) scaling of a "soft" pump, long pulse XeCl laser. (a) pulse length scaling, local efficiencies are shown at 1 and 2 μs. (b) gain length scaling, HCl concentration and FOC optimized for a 2 μs pulse.

References

1.   N. Bloembergen, C.K.N. Patel, et al., "Report to the American Physical Society of the Study Group on Science and Technology of Directed Energy Weapons", Rev. Mod. Phys. 59(3), part II (1987).

2.   Luk, Ting Shan, ed., "SPIE Excimer Lasers and Optics", Vol. 710, pp 710-802 (1986).

3.   K.G. Ibbs and R.M. Osgood, eds., Laser Chemical Processing for Microelectronics (Cambridge, New York, 1989).

4.   The Aurora KrF laser at Los Alamos is such a device. Its 48 beams have put several kilojoules of energy on target in a 5 ns pulse. J. E. Jones et al., "Performance of the Aurora KrF/ICF Laser System"; International Conference on Lasers '89, paper HB.1, New Orleans, LA, 7 Dec 1989.

5.   T.H. Johnson, H.E. Cartland, T.C. Genoni, and A.M. Hunter, J. Appl. Phys. 66(12), 5707 (1989).

6.   The model developed by Kannari, Kimura, and Ewing at Spectra Technology (currently submitted to J. Appl. Phys.) does include vibrational levels greater than v = 2. Their calculations show that the population in these levels is indeed small.

7.   A Boltzmann code has been used to update the electron energy distribution function, and hence the electron rates, at intervals during a simulation. Results obtained with a Boltzmann solver differ only by a few percent from those calculated with the less cumbersome temperature model.

8.   M.J. Berger and S.M. Seltzer, "Tables of Energy Loss and Ranges of Electrons and Positions", NASA Report No. SP-3012 (1964).

9.   W.W. Rigrod, IEEE J. Quantum Electron. QE-14, 377 (1978).

10. A. Sur, A.K. Hui, and J. Tellinghuisen, J. Molec. Spectrosc. 74, 465 (1979); J. Tellinghuisen, J.M. Hoffman, G.C. Tisone, and A.K. Hays, J. Chem. Phys. 64, 2484 (1976).

11. O.L. Bourne and A.J. Alcock, Appl. Phys. Lett. 42, 777 (1983).

12. R.E. Center, J.H. Jacob, M. Rokni, and Z. Rozenberg, Appl. Phys. Lett. 41, 116 (1982).

13. K.Y. Tang and J. Oldenettel, Thermo Electron Technologies (private communication).

14. F. Kannari, Spectra Technology (private communication).

15. W.L. Morgan and M.J. Pound, in the 33rd Annual Gaseous Electronics Conference, Norman, OK, October 1980 (unpublished).

16. By analogy to electron quenching of KrF(B) in M.J. Kushner, D.E. Hanson, and B.I. Schnieder, "Electron Collision Quenching and Energy Deposition Issues in E-beam Pumped KrF Lasers", International Conference on Lasers '89, paper HB.2, New Orleans, LA, 7 Dec 1989.

17. Calculations provided by private communication from D.L. Huestis, SRI International, and S.E. Moody, Spectra Technology.

18. K.Y. Tang and D.L. Huestis, "Calculations of Equilibrium Constants of XeCl $\rightarrow$ Xe + Cl Between 200 and 500K", SRI Technical Report No. MP 79-82, SRI International, Menlo Park, CA (Oct 1979).

19. R.W. Waynant and J.G. Eden, Appl. Phys. Lett. 36, 262 (1980).

20. G.C. Tisone and J.M. Hoffman, IEEE J. Quantum Electron. QE-18, 1008 (1982).

21. L. Litzenberger, R. Slater, and D. Trainor, in Proceedings of the International Conference on Lasers '85, edited by C.P. Wang (STS, McLean, VA, 1986), p. 573.

22. K.Y. Tang and J. Oldenettal, Thermo Electron Technologies (private communication).

23. D.E. Rothe, J.B. West, and M.L. Bhaumik, IEEE J. Quantum Electron. QE-15, 314 (1979).

24. P.J. Hay and T.H. Dunning, Jr., J. Chem. Phys. 69, 2209 (1978).

25. W.D. Kimura, D.R. Guyer, S.E. Moody, J.F. Seamans, and D.H. Ford, Appl. Phys. Lett. 49, 1569 (1986).

Laser-Induced Synthesis of Methoxymethanol (U)

*Robert A. Johnson, Mr. and Ann E. Stanley, Dr.
Research Directorate, AMSMI-RD-RE-QP
U.S. Army Missile Command
Redstone Arsenal, AL 35898-5248

INTRODUCTION

Recently this laboratory has investigated the utilization of lasers for decontamination[1-4]. In the course of these investigations into the dissociation of several chemical agent simulants[1-4], it was noted that the carbon dioxide, infrared, continuous wave laser excitation of methanol produced formaldehyde, which then underwent a consecutive reaction with excess methanol to form a mixture of what was tentatively identified as methoxymethanol in methanol [1,2]. It is interesting that similar investigations[1,2] of the alcohols; ethanol, 1- and 2-propanol showed that although similar species are formed, only methanol was observed to undergo this consecutive reaction.

Since there is considerable interest in methanol as an alternative fuel, a large number of mechanistic studies have appeared. These include pyrolyses[5] shock tube studies[6] and a number of laser probes[7-13]. Practically all of these report the formation of carbon monoxide, hydrogen, formaldehyde, methane, ethene, and acetylene. However, none of these report the consecutive formation of methoxymethanol. The material found in our mixture appears very close to the material described by Lavrovet et al.[14] which is referred to as methoxymethanol (CH_3CH_2OH) or a polymer thereof. Thus, in the open literature, there is no mention that methoxymethanol exists in an isolated form.

Methoxymethanol is indeed an elusive species, as our efforts to utilize conventional techniques to separate and record spectra did not prove successful, since the high interface temperatures normally used, contributed to the decomposition of methoxymethanol. However, in the course of this effort an unconventional gas chromatographic/Fourier transform infrared (GC/FT-IR) technique was developed and used successfully to isolate methoxymethanol and to record relatively high resolution infrared spectra. This technique may be useful to others experiencing similar problems in the identification of reactive species.

Additionally, this development is of military significance, because this is the first reported isolation of methoxymethanol, a material whose properties are totally unknown. The door is wide open for potential military applications, especially in the area of propellants.

EXPERIMENTAL

Samples of methanol were obtained commercially. Methanol, CH_3OH (99.8%), was from Fisher Scientific Company. Deuterated methanol, CD_3OH (99%), was from Stohler/KOR Stable Isotopes Company. The purity of these alcohols was monitored by comparison of gas phase spectra to spectra in the literature[15,16] and gas chromatographic mass spectral analysis. No further purification of the samples was deemed necessary. Standard vacuum techniques were used to handle the samples.

An MKS Baratron Electronic manometer, consisting of a type 222B transducer and a type PDR-5B power supply/digital readout, was used to measure the pressures of the gas phase alcohols in the cell. The laboratory ambient temperatures varied from 20 to 22 °C.

Gas phase reactions were conducted in a stainless steel cell with exterior dimensions of 10.2 x 5.7 x 5.7 cm, and with an optical pathlength of 9.5 cm. Zinc selenide windows were secured to the ends of the cell's long dimensions with 0-rings. The laser beam traversed the 9.5 cm path and the infrared spectra were recorded through the same pathlength.

Infrared laser excitation was provided by a Coherent Radiation model 41 continuous wave (CW), carbon dioxide (CO_2) tunable laser. The laser was operated in a single mode at various selected wavelengths and powers. The selected frequencies and powers were verified using an Optical Engineering CO_2 spectrum analyzer and a Coherent Radiation model 213 water-cooled power meter, respectively. The laser beam was reflected from a flat mirror onto a gold-plated focusing mirror and then into the sample cell.

The sample cell was placed in the laser beam path just behind the focal point, where the beam (about 2 mm in diameter) was slightly diverging. The exterior side of the zinc selenide windows was antireflective coated, which allowed a beam transmittance of approximately 85 percent through the front window. Powers of laser fluences reported here do not correct for this absorption of the window or for the difference in beam diameter at the sample versus at the power meter where the beam is larger. The actual powers at the sample are actually slightly higher than the reported values.

A Mattson Sirius 100 Fourier transform infrared interferometer, with a KBr beamsplitter and a deuterated triglycine sulfate detector, was used to record infrared spectra. Interferograms were transformed after applying a triangular apodization function with an effective spectral resolution of 1.0 cm^{-1}. This resolution is sufficient to permit identification of products with sharp, narrow bands.

Separation was accomplished using two different gas chromatographic systems (GC), both of which were outfitted with the sampling valve described below. Mass spectra were obtained using the Hewlett-Packard model 5890A GC interfaced to a Hewlett-Packard model 5970 mass selective detector (GC/MS). The other system consisted of a Hewlett-Packard model 5890 gas chromatograph with both a flame ionization detector (FID) and a cryogenic valve which permits operation of the chromatographic oven at sub-ambient temperatures. In both systems a Valco 6-port gas sampling valve was used for the vapor-phase sample introduction onto an Alltech, 15 meter, 0.32mm id polydimethylsiloxane capillary column. Deactivated, 0.32mm id, fused silica capillary columns were used for the interface to the lightpipe and from there to the FID detector. A Mattson Instrument, model LYRA lightpipe was used to interface the GC to the Mattson Instrument, Sirius 100 FT-IR (GC/FTIR). The LYRA has a 15.0 cm by 1.0 mm, gold surface lightpipe with KBr windows and a liquid nitrogen cooled mercury-cadmium-telluride detector.

RESULTS AND DISCUSSION

It is known from our earlier studies that the laser-induced reaction of methanol yielded formaldehyde which after formation reacted with excess methanol to produce methoxymethanol[1,2]. The infrared spectra previously recorded using FT-IR spectroscopy was the combined spectra of methanol and methoxymethanol because the formaldehyde formed initially was completely reacted. A typical spectrum of this mixture is shown in Figure 1.
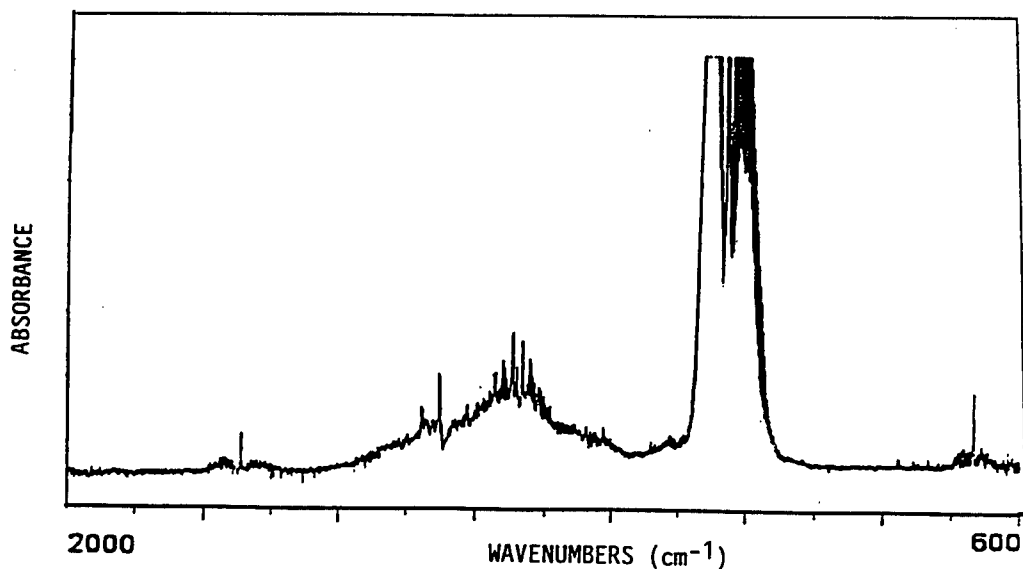
Fig. 1. Infrared (FTIR) spectrum showing the formation of formaldehyde from the irradiation of 50 torr of methanol at $100W/cm^2$ for 0.1 sec.

255

All GC/MS experimental parameters were varied until an optimal separation of the reactant/product mixture was obtained. This optimization yielded the desired separation, including the methoxymethanol. A typical GC/MS total ion chromatogram and mass spectrum is shown in Figure 2. In the chromatogram, Figure 2A, the first eluant is carbon monoxide and formaldehyde. The second eluant is the unreacted methanol. The third eluant is the one suspected to be methoxymethanol. The mass fragmentation pattern for this unknown is shown in Figure 2B. As will be discussed later, this pattern indicates the proposed structure, CH_3OCH_2OH.
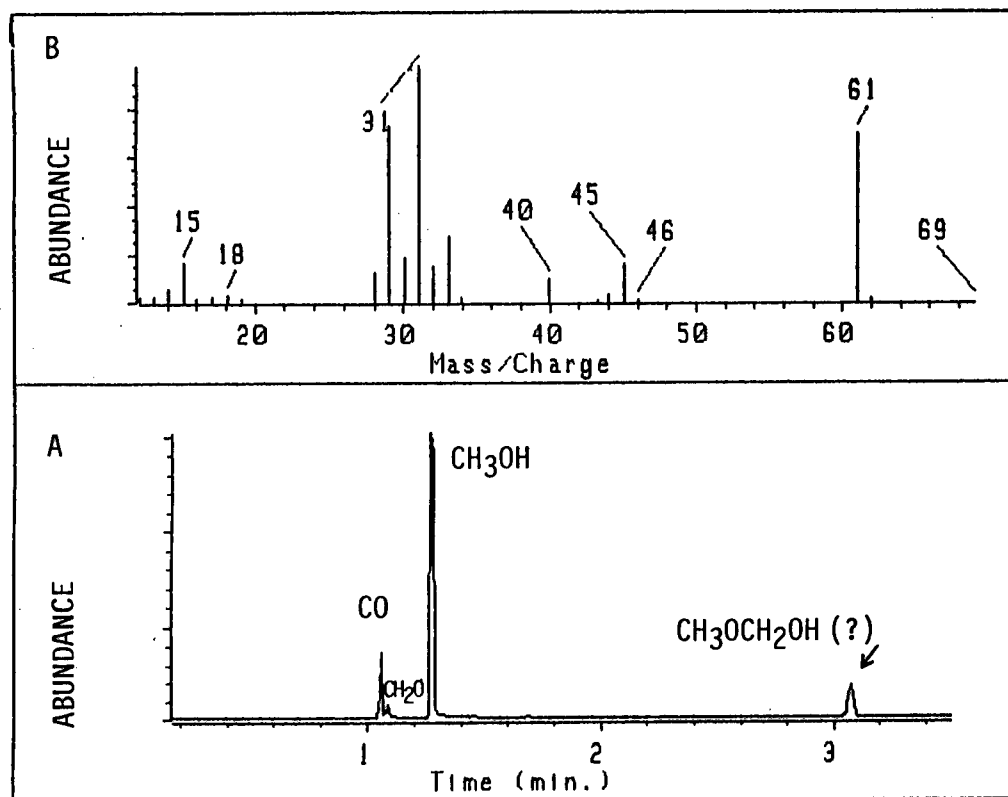
Fig. 2. GC/MS (A) total ion chromatogram (TIC) and (B) mass spectrum of the methoxymethanol formed after the irradiation of methanol (100 torr CH_3OH, 100W/cm^2, 5 sec.)

GC/FT-IR is an established, well known and powerful analytical tool which commonly utilizes a flow-through gold lightpipe as an interface between the GC and a mercury-cadmium-telluride (MCT) detector used for recording infrared spectra of the separated species. Recommended lightpipe operating conditions are temperatures between 100 and 250°C and 4 - 16 cm^{-1} spectral resolution. The optimized chromatographic technique from the GC/MS studies was implemented on the GC/FTIR system using conventional parameters. The Gram-Schmidt reconstructed chromatogram showed three separate peaks, one each for formaldehyde, methanol, and methoxymethanol, in that order of elution. Upon investigation of the infrared spectrum of that second peak, formaldehyde (1746 cm^{-1}) and methanol (1033 cm^{-1}) were identified, suggesting the decomposition of the methoxymethanol. A typical spectrum is shown in Figure 3.
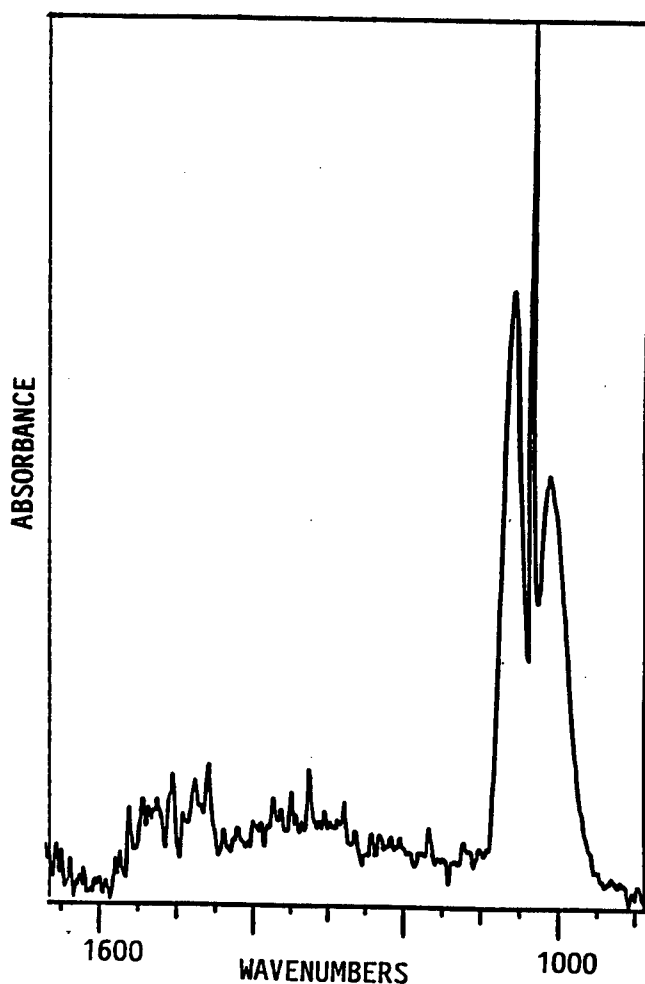
Fig. 3. Infrared (GC/FTIR) spectrum showing the decomposition of the isolated, decomposed methoxymethanol at a lightpipe temperature of 125°C.

In our situation the high interface temperature was contributing to the decomposition of an unstable reaction product, thus the possibility of operating at significantly lower temperatures was investigated. The lightpipe was gradually cooled to room temperature and the cryogenic capability of the chromatograph was used to lower the oven temperature to 10 degrees Celsius. At this point, the infrared spectrum showed that some undecomposed methoxymethanol remained. The lightpipe compartment was modified to allow introduction of dry ice which cooled the lightpipe and transfer line while also maintaining the nitrogen purge in the interface compartment. The temperature of the lightpipe and interface column was approximately -20 degrees Celsius. After a few trials at this temperature, only methoxymethanol (1046, 1021 cm^{-1}, etc.) was found in the third peak, as shown in Figure 4. This successful trial was repeated and refined until the infrared spectrum of methoxymethanol was recorded at a resolution of 1.0 cm^{-1}. Figures 3 and 4 show the 4 cm^{-1} resolution spectra of two separate trials, one with the lightpipe at an elevated temperature and the other at a sub-ambient temperature.
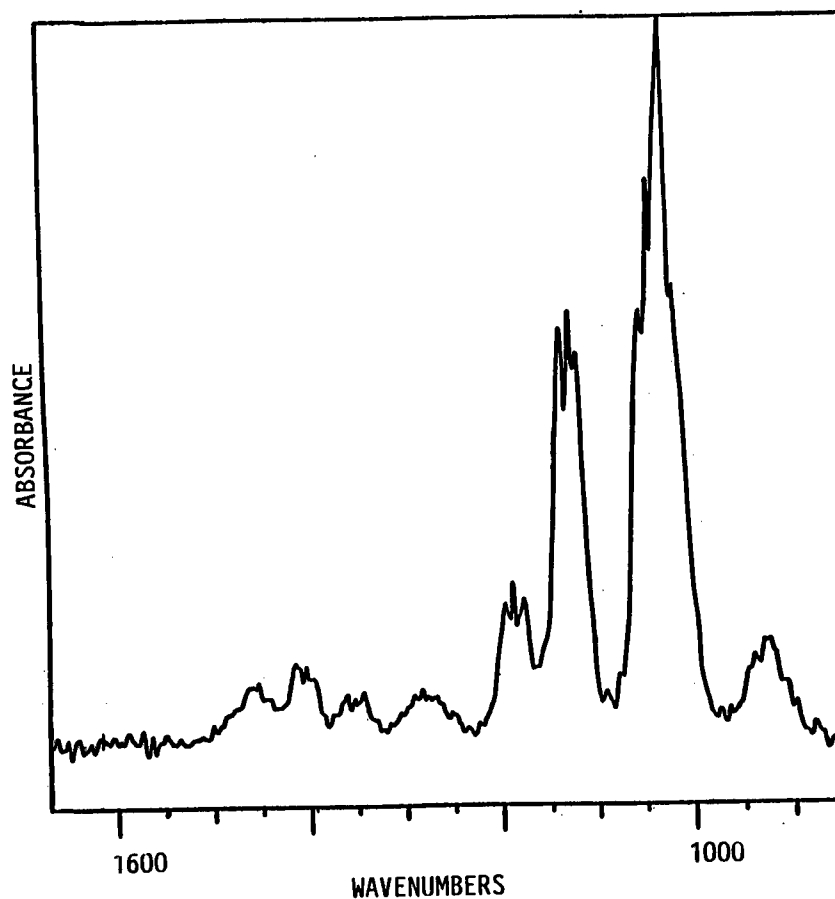
Fig. 4.   Infrared (GC/FTIR) spectrum showing the decomposition of the isolated, decomposed methoxymethanol at a lightpipe temperature of -16°C.

After this successful analysis, the same techniques were used to investigate one species of deuterated methanol, CD_3OH. A typical TIC and mass spectrum for the deuterated methoxymethanol is shown in Figure 5. Analogous to the results obtained for ordinary methanol, the first eluant is carbon monoxide and formaldehyde. The second eluant is unreacted methanol, CD_3OH. The last eluant is methoxymethanol. The mass fragmentation pattern of that eluant shown in Figure 5B supports the methoxymethanol structure. As discussed later, it is clear that at least CD_3OCD_2OH is present.
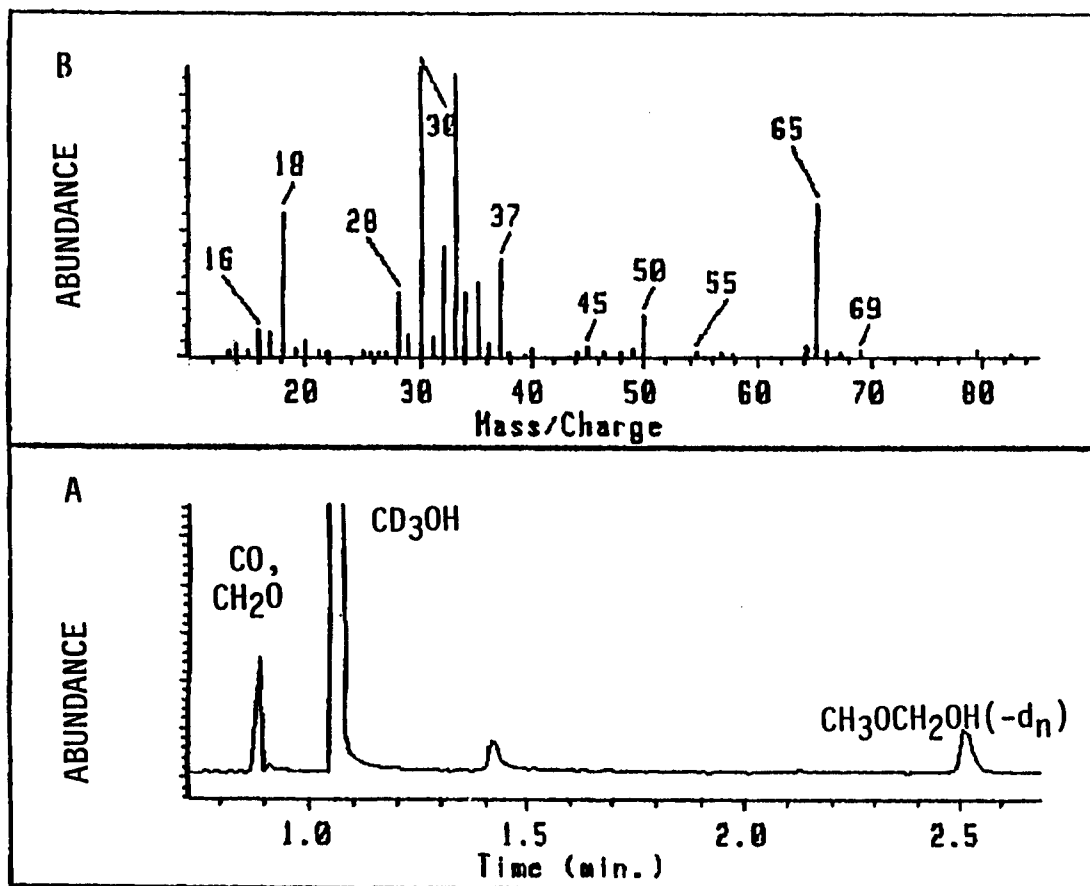
Fig. 5. GC/MS (A) total ion chromatogram (TIC) and (B) mass spectrum of the deuterated methoxymethanol ($-d_n$) formed after the irradiation of methanol-d_3, (100 torr, CD_3OH, $100W/cm^2$, 5 sec.)

259

The same high temperature GC/FTIR technique was used in the analysis of the mixtures produced after the irradiation of 100 torr of CD_3OH, at a power of $100W/cm^2$ for 5 seconds. A Gram-Schmidt reconstructed chromatogram also yielded three separate peaks for the formaldehyde species, unreacted methanol ($-d_3$) and the methoxymethanol species. Analysis of the infrared spectrum of the formaldehyde peak showed that CD_2O, (2056 cm^{-1}, 1700), CDHO (1400 cm^{-1}, 2844, $2120, 1723$) and CH_2O (2780 cm^{-1}, 1750) all are present. Analysis of the methoxymethanol ($-d_n$) spectrum, Figure 6, showed, as expected, that the methoxymethanol was decomposing to formaldehyde ($-d_0$, $-d_1$, $-d_2$) and methanol ($-d_3$).
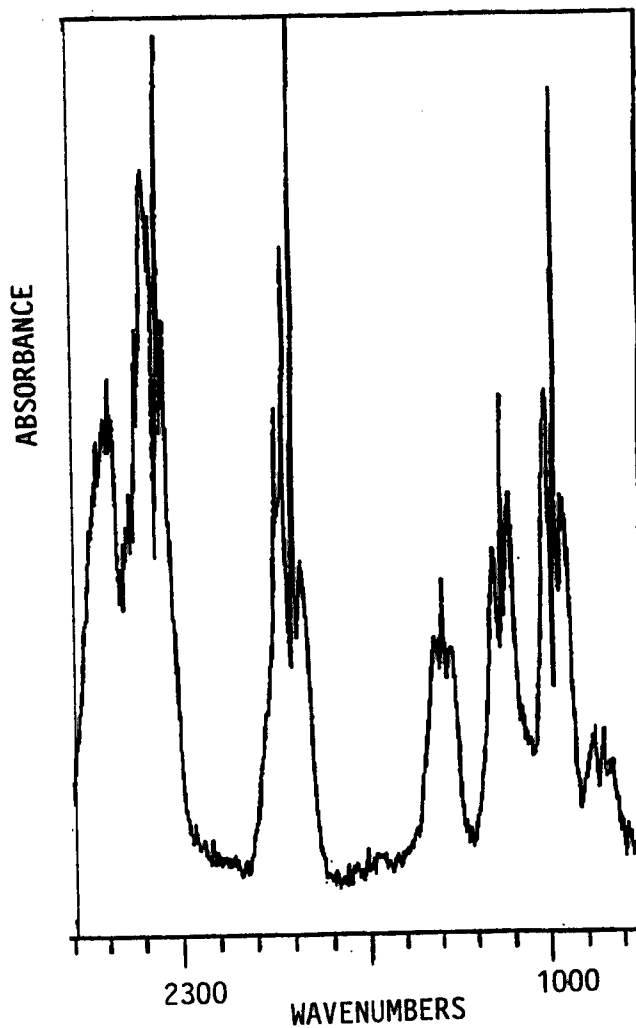
Fig. 6.   Infrared (GC/FTIR) spectrum showing the decomposition of the isolated, decomposed methoxymethanol ($-d_n$) at an elevated lightpipe temperature (100 torr, CD_3OH, $100W/cm^2$, 5 sec.)

As before, the oven temperature, interface and lightpipe temperatures were too high. Thus, the cryogenic techniques described earlier were implemented on the GC/FTIR system, resulting in an oven temperature of 10°C and a subambient lightpipe temperature. The analysis of the spectrum, Figure 7, of the methoxymethanol shows no decomposition. Particularly dramatic is the <u>absence</u> of all formaldehyde species.
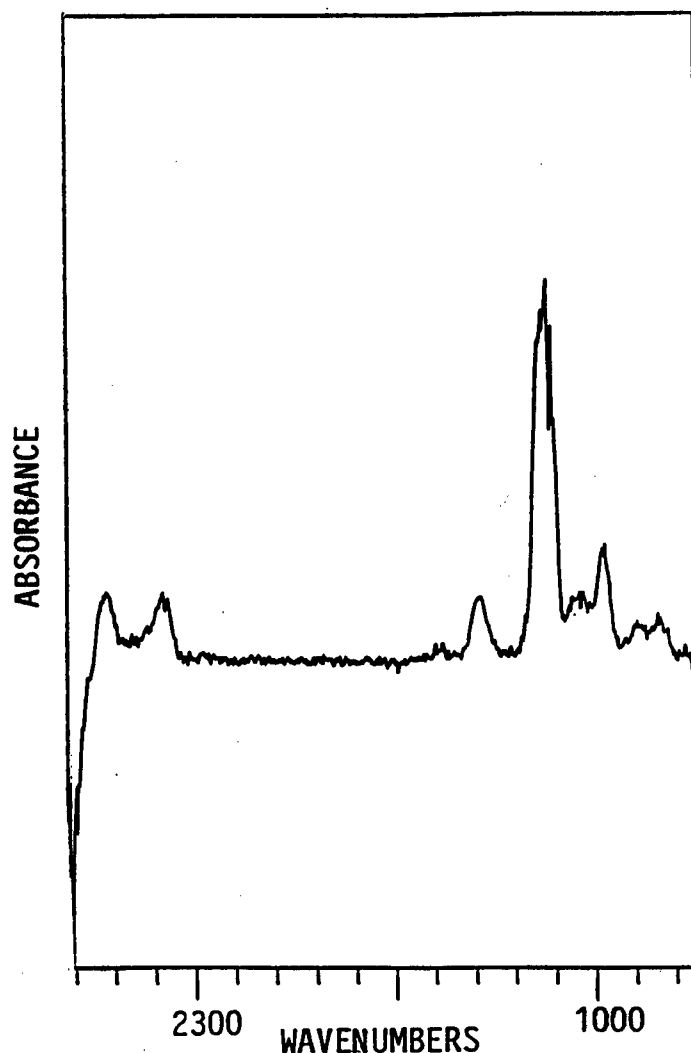
Fig. 7.   Infrared (GC/FTIR) spectrum showing the decomposition of the isolated, nondecomposed methoxymethanol ($-d_n$) at a lightpipe temperature of 6°C (100 torr, CD_3OH, 100W/cm², 5 sec.).

CONCLUSION

The mass fragmentation patterns found in the GC/MS analysis are presented in Table 1. These patterns clearly support the methoxymethanol structure, CH_3OCH_2OH. The assignment presented is not the exclusive interpretation but is a very strong argument for methoxymethanol. In the mass spectrum some formaldehyde and methanol species is likely present because of the high (150°C) interface temperature used in GC/MS analysis. However, it is clear that the predominant species is the higher molecular weight species methoxymethanol.

From the GC/FTIR analysis it is clear that no formaldehyde species is present when the lightpipe temperature is lowered. The frequencies noted in those GC/FTIR as belonging distinctively to methoxymethanol (d_0 and d_n) are presented in Table 2.

Table 1. Assignment of Mass Spectral Fragmentation Patterns for Methoxymethanol (-d_0 and -d_n)

METHOXYMETHANOL (-d_0)		METHOXYMETHANOL (-d_n)	
Mass	Fragment	Mass	Fragment
61	CH_3OCH_2O	65	CD_3OCD_2O
46	OCH_2O	50	CD_3OCD_2
45	CH_3OCH_2	45	CH_3OCH_2
33	CH_3OH_2	37	CD_3ODH
33	CH_3OH_2	35	CD_2HOH_2
32	O_2	34	CD_3O, CD_2OD
31	CH_3O, CH_2OH	33	CD_3OH
30	CH_3O	32	O_2
29	HCO	30	CH_2O
		28	CO

Table 2. Infrared Frequencies* (cm^{-1}) Assigned to Methoxymethanol (-d_0 and -d_n)

Frequencies CH_3OCH_2OH (-d_0)	Frequencies CH_3OCH_2OH (-d_n)
3695	3671
1450	3648
1410	2223
1350	2085
1189	1147
1129	1124
1052	1112
1046	979
1020	898
930	845

* Only Q branch frequencies are noted

The spectra for the light methoxymethanol are easily understood; however, in the deuterated species, there are a few frequencies which could be assigned to CH_3OD or CD_3OH. Considering both the GC/MS and GC/FTIR spectra, it is certain that at least CD_3OCD_2OH is present. In the GC/MS spectra, the mass is correct for the loss of one hydroxy hydrogen. This fits well with the 3671 cm^{-1} or 3648 cm^{-1} OH frequencies. The infrared bands at 2223 and 2085 are obviously CD stretching frequencies. However, since, on decomposition, all three forms of formaldehyde are formed, CH_2O, $CHDO$ and CD_2O, this is an indication that CD_3OCH_2OH, CD_3OHDOH, CD_3OCH_2OH, or another species may be present.

The isolation and characteristics of methoxymethanol are being explored. The utilization of methoxymethanol as a precursor to a powerful, smokeless propellant is very promising and is being attempted.

Additionally, a very useful analytical technique for the separation and identification of a reactive species has been developed. The capabilities of GC/FTIR analysis extend far beyond the conventional recommended conditions with respect to the lightpipe temperature and spectral resolution.

This low temperature lightpipe technique has afforded this laboratory the ability to record the infrared spectrum of an elusive, unstable reaction product. The cryogenic chromatograph oven and this GC/FT-IR technique are quite complementary and should be useful for many applications. The analysis of the higher resolution spectra of the normal and deuterated species are underway and will be the subject of another article.

REFERENCES

1. R. A. Johnson and A. E. Stanley, "Laser-Induced Destruction of Some Chemical Agent Simulants: The Alcohols," Technical Report RD-RE-87-2 (U.S. Army Missile Command, Redstone Arsenal, AL, February 1987).

2. R. A. Johnson and A. E. Stanley,"Laser-Induced Photochemical Reactions of Methanol, Ethanol, 1-Propanol, and 2-Propanol, " Appl. Spectrosc. 42, 1268 (1988).

3. A. E. Stanley, MICOM Technical Report TR-RD-RE-86-13 (U.S. Army Missile Command, Redstone Arsenal, AL, 1986).

4. A. E. Stanley, MICOM Technical Report TR-RD-RE-86-15 (U.S. Army Missile Command, Redstone Arsenal, AL, 1986).

5. J. M. Fletcher, Proc. R. Soc. (London), Ser. A, 147, 119 (1934).

6. D. S. Bomses, S. Dougal, and R. L. Woodes, J. Phys. Chem. 90, 2640 (1986) and references therein.

7. S. E. Bialkowski and W. A. Guillory, J. Chem. Phys. 67, 2061 (1977).

8. R. V. Ambartzumian, N. V. Chekalin, V. S. Letokhov and E. A. Ryabov, Chem. Phys. Letters 36(3), 301 (1975).

9. S. W. Mayer, M. A. Kwok, R. W. F. Gross and D. J. Spencer, Appl. Phys. Letters 17(12), 516 (1970).

10. J. D. Campbell, M. H. Yu and C. Wittig, Appl. Phys. Letters 32(7), 413 (1978).

11. S. E. Bialkowski and W. A. Guillory, J. Chem. Phys. 68(8), 3339 (1978).

12. R. D. McAlpine, D. K. Evans and F. K. McClusky, Chem. Phys. 39, 263 (1979).

13. R. Bhatnagar and P. E. Dyer, Chem. Phys. 61(2), 339 (1979).

14. O. A. Lavrova, Z. A. Matreeva, T. M. Lesteva, and B. I. Pantukh, Russian J. of Phys. Chem. 49, 374 (1975).

15. M. Falk and E. Whalley, J. Chem. Phys. 34(5), 1554 (1961).

16. A. Borden and E. F. Barker, J. Chem. Phys. 6 , 553 (1938).

How Much Soldier Quality?
Estimating Cost-Effective Recruit Selection Policy
for a Smaller Army (U)

Cyril E. Kearl
Roy D. Nord
Manpower and Personnel Policy Research Group
US Army Research Institute
Alexandria, VA 22333-5600

I. Introduction

Anticipated force reductions have dramatically reduced annual recruiting and accession targets. By selectively reducing recruiting for particular types of enlistees, The United States Army Recruiting Command (USAREC) can enrich the manpower quality mix of the accession cohort, thus increasing soldier performance and reducing first-term attrition. On the other hand, since the marginal cost of a high-quality recruit is roughly $20,000 more than that for low quality, a policy of across-the board reductions in recruiting would allow the Army to make larger reductions in its recruiting budget.

Army manpower policies establish specific quality mix targets as well as total annual accession missions. Since the beginning of FY90, overall Army accession targets have been slashed from more than 110,000 to 92,000 -a cut of more than 15%. At the same time, missions for high quality recruits remain approximately constant, increasing high quality content in the accession cohort by more than 13%. Is this increase in quality content justified? The answer depends on the value of increased soldier quality relative to its costs. If the net payoff to higher quality soldiers exceeds that of their lower quality substitutes, then a larger recruiting budget than that required to obtain minimally qualified manpower is justified.

In order to make the required comparison, it is necessary to know (or be able to estimate) several key relationships:
First, what is the relationship between recruit "quality" (defined in terms of scores on the Armed Forces Qualification Test (AFQT) and educational attainment) and soldier performance?
Second, what is the value or payoff to different levels of soldier performance?
Third, what are the costs of recruiting and training soldiers, and how do these costs vary with the quality mix?

While there has been considerable research on each of these questions individually, there has been no integrated analysis of the kind required to answer the question we address here -- that is: What is the most cost-effective mix of recruit quality, and how should that mix vary as force size changes? Our results support the Army Recruiting Command's decision to increase the manpower quality mixture. We find that smaller accession cohorts make higher manpower quality mixes more cost effective. Our approach also provides estimates of recruiting costs for a particular size accession cohort with a given quality mix. This permits us to estimate the appropriate recruiting budget size for a given accession cohort (and particularly for the cost effective cohort). Given the expected

reductions in recruiting budgets over the next five years, these estimates may be helpful in developing budget plans that allow the maintenance of a cost effective quality mix as force structure changes.

Our approach is first to estimate the value of net performance payoffs for potential recruits. We then use a linear program to assign recruits to MOS so that the most cost effective job matches are found. In Section II, we outline our methodology for estimating performance payoffs. Recruiting and attrition costs are addressed in sections III and IV, respectively.

In order to make the optimization problem practically feasible, it is necessary to group recruits and jobs into clusters. An outline of our methods for clustering and the specification of the linear program is presented in Section V.

Section VI describes the optimization model. Section VII concludes the paper by presenting some preliminary results, and outlining the ongoing analysis currently being carried out with the model.

II. Estimating Net Performance

At the heart of this work is the valuation of net performance. Estimating the value of net performance requires four separate sets of estimations: gross performance; recruiting costs; attrition costs; and the probability of attrition. Since it is not possible within the scope of our research to estimate models for each of these components we have relied upon previous research for estimates of much of the gross performance valuation, recruiting costs, and for some of the training costs. In this section we briefly describe the models used and the rationale for their selection.

Soldier Performance. During the last decade, the U.S. Army has accessed between 110,000 and 130,000 nonprior service recruits each year. A substantial proportion of the Army's manpower is provided by soldiers serving in their first enlistment term. As a result, the performance of these soldiers has a significant effect on the overall readiness of the Army. In addition, the performance of first term enlistees determine the quality of future Army noncommissioned officers because of the Army practice of relying on internal lines of promotion. The Army makes a considerable investment administering tests that accurately predict performance. The Armed Services Vocational Aptitude Battery (ASVAB), given to applicants prior to accession, is used to generate a set of nine "aptitude area composite" (AA) scores that are used to predict job-specific performance. Each military occupational specialty (MOS), requires a minimum score on at least one AA composite. MOS. An additional composite of ASVAB scores is used to construct the Armed Forces Qualifying Test (AFQT) score, the results of which are used as the primary entry screen for the Army.

There is a large body of research that provides convincing evidence that soldier quality is strongly related to training success, job performance, and the probability of completing the enlistment term. Unfortunately, these studies have dealt only with a few isolated specialties, and do not provide a way to make direct tradeoffs against dollar-valued recruiting and training costs. The best existing measure of performance that extends to a wide range of MOS is the Skill Qualification Test (SQT)[1]. The SQT is a MOS specific written test. It is used to test proficiency for more than 80% of the jobs for which first term soldiers enlist. SQT test scores range between 0 and 100 with scores above 60 considered passing.

There are difficulties using SQT as a measure of performance. Most seriously, the distribution of SQT scores varies dramatically across MOS and over time. This limits the usefulness

of SQT in making performance comparisons -- a critical factor when using SQT to estimate relative performance payoffs from recruit assignment[2]. In addition, the SQT is a written test, is an extremely limited measure of performance. Finally, the SQT is not administered for approximately 20% of all MOS.

An alternative approach to measuring expected performance is to rely on performance predictors such as the Armed Forces Qualification Test (AFQT) and MOS-specific Aptitude Area (AA) scores. Performance predictors have an advantage of being validated over a much broader range of performance measures including SQT. Both AFQT and AA scores are positively related to performance of essential Army tasks for specific jobs[3] Furthermore, both tests are correlated with performance on the SQT[4]. However, Frances Grafton and David Horne found that the AA is the single, most significant, predictor of SQT performance for nearly every MOS[5]. D. H. McLaughlin, et al. have found AA scores to be the best predictor of training success[6]. MOSAA has also been found to be a superior predictor of Core technical performance and general soldiering skills[7]. Consequently, we have adopted AA score as our predictor of performance differences. This implies:

$$\sigma_{\text{Performance}} = \alpha(\sigma_{\text{AA}}), \tag{1}$$

where $\sigma_{\text{performance}}$ is the standard deviation of job performance, σ_{AA} is the standard deviation of AA scores, and α is the validity of AA as a predictor of performance.

Joseph Zeidner and Cecil Johnson have analyzed a number of studies relating MOSAA to performance. Using Empirical Bayes techniques, they estimate AA validities for nine "families" of similar jobs. These estimates range between .55 and .60, with a mean of .58[8].

It is still necessary to determine a market value for performance. We assume that this value is positively related to performance:

$$\sigma_{\text{value}} = \beta\ (\sigma_{\text{Performance}}), \quad \beta > 0, \tag{2}$$

where σ_{value} is the standard deviation of the distribution of dollar-values performance.

Since there is no market valuation of military service, it is not feasible to determine the value of performance based on the value of output. In addition, differences in earnings which are expected to reflect differences in the value of performance are almost nonexistent during the first term of enlistment. However, there has been considerable research in the civilian sector on the evaluation of both performance predictors and job performance itself in terms of dollar value as measured either by compensation or variations in output. While the relevance of this research to the evaluation of soldier performance may at first seem questionable, we argue that (a) it is not unreasonable to assume that the rate of substitution between labor quality and other inputs in the military is similar to that in the economy at large; and (b) that, since a volunteer Army must compete with the private sector to attract its manpower, competitive forces insure that the value to society of the output produced by that manpower must be approximately equal to its price -- i.e., to average military compensation. By this logic, it is possible to use the results of performance valuation research in the civilian sector to calibrate soldier performance in terms that allow for tradeoffs between recruiting costs and performance value.

One source for estimates of the returns to performance is research measuring the distribution of earnings for wage earners in the United States: both within job classifications and

for the economy as a whole. Frank L. Schmidt and John E. Hunter have summarized both types of research[9]. They estimate the standard deviation of the distribution of earnings is between 40% and 60% of mean wages. If the measure of job performance is standardized to have mean 0 and standard deviation 1, this implies $\beta = .4\text{-}.6$. Using these estimates of β, combined with an assumed mean validity of .58, means that the average soldier with an AA score one standard deviation above average would produce output valued of between 23% and 34% more than average. We have adopted the most conservative estimate in this range, 23%.

Having developed an approach to determine the value of performance based on MOSAA, our procedure is to compute this value as a multiple of mean first term earnings. We base first term earnings on the present value of regular military compensation of those entering active Army service during FY1984[10]. In all, our sample included 131,937 men and women. Regular military compensation (RMC) includes both basic pay, and "in kind" compensation - basic allowance for quarters, basic allowance for subsistence, and the federal tax advantage (since the "in kind" benefits are not taxable)[11]. Since RMC varies by pay grade and length of service, we calculated mean time to promotion to construct an average earnings profile. As mentioned above, promotions through E-3 appear to be fairly automatic so that there is little variation around mean promotion times for these pay grades[12]. The average earnings profile was constructed over the thirty-nine months, the average term of a first enlistment. The present value of this earnings profile was computed assuming an 8% annual discount rate. To insure comparability of the value of performance to training and recruiting costs, our calculations were in 1989 dollars. In addition to promotions an longevity increases, soldiers also receive annual wage increases. Estimates of these are not included in our calculations since they vary with actual and expected inflation and act as "cost-of-living" increases just compensating for increases in inflation which is also omitted from our calculations.

We estimate the present value of average first enlistment term performance to be $46,000. This represents the expected value of individual performance from recruits with average AA scores. The expected value of performance for those with AA one standard deviation above average is $56,580. The value of performance will depend on the characteristics of the individual and of the particular job under consideration:

$$v_{ik} = \alpha\beta(AA_{ik})$$

III. Estimating Recruiting Costs

There has been considerable research on supply functions for high quality enlistments. Estimates of supply functions for quality markets model the effects of recruiting incentives, demographic conditions, and economic conditions almost exclusively for the male labor supply. However, many of these estimates are based on physical quantities, so that additional calculations are needed to obtain cost estimates of the effects. Consequently, in the short run recruiting costs are known only imprecisely for many of the recruiting resources, making it difficult to derive accurate marginal recruiting cost estimates. Furthermore, marginal cost estimates for different recruiting resources exhibit more variation than is consistent with efficiency. Determining the marginal cost of obtaining an additional enlistment could vary significantly depending on the type of recruiting incentives employed. However, it is unreasonable to assume that these significant differences persist over long periods of time without a reallocation of recruiting expenditures toward the recruiting resources with the lowest marginal costs. One recruiting incentive for which the marginal cost is relatively well defined is regular military compensation because it is dollar denominated. This makes the coefficient of earnings on enlistments a likely choice as a single estimate of marginal costs. We follow Armor, et al. in adopting two assumptions about the supply

elasticity for the quality labor market. Our high estimate is 1.0 and our lower estimate is .5. Since the U.S. Army is the single largest employer in the quality youth labor market, its manpower needs have considerable effect on this market. Consequently, marginal costs are estimated by $MC = W(1 + 1/\epsilon_{supply})$[13] and $\epsilon_{MC} = 1/\epsilon_{supply}$.

RMC is determined by act of Congress[14] so that for a particular compensation level, w_0, there will be a supply of recruits, Q_0. For enlistment requirements up to Q_0, recruiting costs are constant, equal to the processing costs necessary for testing and screening. For manpower needs in excess of Q_0, additional recruiting incentives will be necessary to attract the desired manpower and the supply curve will be upward sloping as recruiting costs increase with the number of recruits required. To calculate Q_0, we have estimated average recruiting costs for high quality enlistments for FY1986 (excluding wages). For a given Q_0 and ϵ_{supply} marginal recruiting costs can be calculated for each possible enlistment demand. Q_0 occurs where marginal and average recruiting costs are approximately equal. For our optimization we calculated that $Q_0 = 17,000$ with estimated processing costs of \$3,068 per recruit in 1986.

The point at which marginal costs begin to rise depends on economic conditions. Changes in unemployment can be expected to shift the supply curve and consequently to shift Q_0 in response. We use an unemployment elasticity of 0.8 and adjust Q_0 based on the percentage change in unemployment since 1986. For example, the decline in unemployment from 7.2% in FY1986 to 5.3% in FY1989 represents a 32% drop in unemployment and reduces Q_0 by 24% to 12,920. With tighter labor markets, marginal recruiting costs begin increasing sooner and reach higher levels for a given manpower requirement.

In contrast to the quality male labor market, there has been little empirical research to estimate the supply functions for the non-quality labor markets. Previous research has taken it for granted that non-quality recruiting populations are sufficient to support Army demand at current wages. Where costs estimates for recruiting non-quality markets are required, they are assumed to be equal to processing costs[15]. Nevertheless, there is substantial evidence that recruiting costs are not constant for these groups. The problem of increasing marginal costs of the non-quality groups was particularly evident during FY1989. When the Recruiting Command reluctantly increased the proportion of non-quality recruits it enlists, it found that recruiting bonuses (normally reserved only for quality enlistments) were necessary to attract sufficient numbers from some of the superior non-quality markets. In addition, two segments of the non-quality labor market were utilized to meet manpower requirements that are normally closed to enlistment (non-high school graduates in AFQT CAT IIIB (see Table 1) graduates with AFQT scores below 25). This not only expanded the non-quality market but it also directly increased recruiting costs by increasing the number of applicants that were rejected for not meeting minimum standards for any available MOS increasing the recruiting time necessary to find acceptable applicants.[16]

The non-quality market is actually composed of several market segments with systematic differences in productivity and recruiting costs. We assume that these markets share the same more responsive supply curve with respect to recruiting incentives. Consequently we use $\epsilon_{supply} = 2.0$ for estimating the recruiting costs in the non-quality market. On the other hand, Q_0 will depend on market size and attitudes toward military service which vary across market segments. We calculate Q_{0i} using National Longitudinal Survey data for the ith non-quality market segment. Both the population and the attitudes toward military service are compared to those in the quality male market segment and Q_0 is adjusted to reflect these differences:

$$Q_{0i} = (P_{0i}/P_0)*(A_{0i}/A_0)*Q_0 \qquad (3)$$

where P_{0i} represents the population in the ith submarket P_0 represents the population in the male quality market; A_{0i} represents the fraction responding positively toward military service in the i_{th} market; and A_0 represents the fraction responding positively toward military service in the male quality market.

Table 1
Army Recruiting Mission Boxes

Name	Description
GMA	Graduate Males AFQT Cat I-IIIA
GFA	Graduate Females AFQT Cat I-IIIA
GMB	Graduate Males AFQT Cat IIIB
GFB	Graduate Females AFQT Cat IIIB
GM4	Graduate Males AFQT Cat IV
NMA	Non-graduate Males AFQT Cat I-IIIA

A similar approach is adopted to calculate Q_0 for both the female quality and non-quality market segments. In all separate supply and marginal costs curves were calculated for six markets, defined as shown in . Table 2 summarizes the estimated Q_0 and assumed elasticity of supply for each of these market segments.

Table 2
Submarket Characteristics

Mission Box	Q_0	Elasticity
GMA	17,000	1
GFA	3,060	1
GMB	5,714	2
GFB	1,461	2
GM4	3,019	2
NMA	1,502	2

Overall Recruiting Costs in the ith recruiting market (at a given time, t) is determined by the parameters of the supply function and the number of recruits from that market segment:

$$\phi_i = \phi(Q_{0i}, \epsilon_i, x_i) \tag{4}$$

IV. Attrition Costs

Unlike civilian employment, recruits joining the Army are required to sign an enlistment contract legally binding them to serve a specified amount of time. Nevertheless, separation prior to the completion of the enlistment term is a frequent occurrence. Attrition imposes costs on the Army. In addition to administrative costs, attrition is frequently related to disciplinary or medical problems both of which increase costs. There are also efficiency costs. Since attrition is largely unanticipated, units may be forced to operate below desired strength until suitable replacements can be enlisted and trained.

unanticipated, units may be forced to operate below desired strength until suitable replacements can be enlisted and trained.

Training represents a significant component of the first enlistment term. Virtually all recruits (not previously in the Army) receive general orientation to the Army in the form of Basic Training and initial skill training for their particular MOS. The length of time necessary to complete training varies depending on the complexity of the skills being taught. Training can be as short as 15 weeks or can take as long as a year. MOS which need long training periods also require longer enlistment terms to ensure that the training costs can be recovered during the enlistment term. Training costs have two primary components: the costs of the facilities and equipment necessary to conduct the training and the wages and salaries paid to the trainees during training while the enlistees are not productive. The data used to compute training costs was extracted from the Resource Management Office of the U.S. Army Training and Doctrine Command (TRADOC) based on the ATRM-159 report[17].

Unfortunately, there is insufficient data to estimate the costs of efficiency loss, medical costs, and disciplinary costs that are incurred as a result of attrition. Consequently, we assume that the cost of an attrition (regardless of when it occurs) is equal to a multiple of the cost of training a replacement.

We estimate the probability of attrition using a logistic regression as a function of the market segment characteristics, MOSAA, term of enlistment and race. Because attrition rates vary across MOS, separate attrition equations are estimated for each MOS cluster using a logistic regression and cohort data from 1982-1984. (Parameter estimates are available upon request.) Expected attrition rates are then calculated, based on the parameter estimates and the mean characteristics of the recruiting population in each market segment and expected attrition costs are computed:

$$\cap_{ijk} = \Omega_{ijk} \; (3\tau_k) \tag{5}$$

where Ω_{ijk} represents the probability of attrition for the jth individual in the ith market segment in the kth MOS; τ_j represents training costs for the kth MOS; and $\cap$ represents the expected attrition cost. Computational constraints required that we limit the number of MOS to which recruits could be assigned.

The Net Value of Performance. Net performance, π, can be derived as a combination of performance value, recruiting costs, and attrition costs. However, because of the sequential way in which recruiting costs are incurred prior to any performance payoffs or attrition costs, it is convenient to exclude recruiting costs from the definition of "net performance value". Total value can be expressed as net performance value less recruiting costs:

$$\pi_{ijk} = V_{ik} - \cap_{ijk} \tag{6}$$

and

$$V_{ijk} = \pi_{ijk} - \phi_i \tag{7}$$

Clustering Supplies and Demands

The job performance obtained from a given recruit population depends not only on the quality mix in that population, but also on the way the pool of job-specific aptitudes residing in that population is allocated across jobs. A given recruit population will yield substantially higher performance if it is optimally allocated across jobs than if job assignment is random. In general, more efficient job assignment procedures make quality more valuable. Thus, in order to determine the cost-effectiveness of a given recruiting policy the job assignment process must be modeled. If this process were modeled directly, it would require a simulation involving more than 18 million variables and 94,500 constraints.

Not only is such a problem not feasible given current research resources, but the solution it produced would also be misleading. In effect this specification relies on information that is, in practice, not available at the time the selection decision is made. The Army targets its recruiting efforts in terms of "mission boxes" defined as shown in Table 1. Information on the job-specific aptitudes of applicants is not available until most of the costs associated with recruiting a given applicant have already been incurred. The primary use of aptitude area scores in the current system is to manage the assignment process (using job-specific minimum eligibility standards -- "cut scores"). A simulated assignment process that relies on individual-level information about aptitudes and job availability would be "too efficient" -- that is, it would yield overly optimistic estimates of performance payoffs, and thus an overestimate of the cost effective recruiting budget.

Finally, the use of individual data might yield the maximum payoff from the specific population assigned, but it would not necessarily produce the best estimate of the expected payoff from a different population with the same representation across mission boxes. Because the primary objective in this exercise is to identify the most cost-effective selection strategy, and that strategy is defined at the level of mission box rather than individuals, the grouping of applicants into clusters is desirable from an analytic as well as a practical standpoint.

Supply Clusters: Recruit supplies are grouped on two dimensions -- by mission box, and by "eligibility profile". The general objectives of the clustering are to (a) define categories of supply that are homogeneous with respect to both recruiting costs and job eligibility; and (b) to reduce the optimization problem to manageable proportions. We employ a three-stage clustering procedure to do this:

In the first stage, we select a large random sample from an actual accession cohort (actual accessions, rather than a sample of the general youth population, are used to avoid having to model the complex processes by which individuals become applicants). For this analysis we used a 50% sample of the 1986 accession cohort. For each individual in the sample, we construct an "eligibility profile", a vector of zeros and ones, indicating the specific jobs for which the individual is qualified. For the 107 MOS included in our sample, there are 33 possible combinations of required AA composite and minimum score, thus the "eligibility profile" contains 33 elements. We then create an initial set of clusters made up of individuals with identical profiles. While this approach has the potential of yielding as many clusters as there are individuals in the sample, these is considerable homogeneity in the distribution -- there were a total of 4967 eligibility profiles among our sample. The final step in the first stage is to subdivide the eligibility profiles by mission box, this yielded an inital total of 9473 supply clusters.

The second stage of the procedure uses a statistical clustering technique to aid in the selection of a subset of the initial clusters to serve as "seeds" for the final stage. The purpose of this exercise is to insure that the selected seeds "span" the range of different eligibility profiles. To accomplish this we used the SA procedure FASTCLUS to generate approximately fifty "super-clusters" using the mean AA scores for each first-stage cluster as the distance criterion.

The final clusters were generated by first generating a set of between 30 and 80 seed clusters for each mission box, and then combining each remaining "candidate" clusters with the "closest" seed, where closeness is defined as the minimum of the sum of positive differences between the minimum qualifying scores of the seed and the candidate cluster. The seeds were defined by selecting the largest first-stage supply group from each mission box within each of the 50 second-stage clusters, plus any other supply groups with at least 200 members. (Not all mission boxes were represented in every cluster.) The distance criterion was imposed by summing the positive differences between a candidate and each seed for the same mission box, and taking the minimum of these distances. Negative distances -- implying that the candidate was unqualified for at least one MOS for which the seed cluster was eligible were penalized as follows: any deviation greater than or equal to five points disqualified the match. If no strictly positive sum was found for any seed, seeds with negative distances of up to 4 points in one or more aptitude areas were considered.

The final result of this approach was a set of 147 different aptitude profiles that, combined with mission box, yielded a total of 341 supply clusters. The mean absolute difference between the eligibility vectors of the final clusters and those originally defined averaged .75 points for each aptitude area.

MOS Clusters. The primary objectives for the MOS clustering were to insure homogeneity within clusters with respect to performance, attrition, and eligibility. An additional criterion was to obtain clusters with sufficient numbers of accessions to allow the attrition prediction equations to be reliably estimated. Our approach to this was simple. We grouped MOS on the basis of Career Management Field (CMF), the primary aptitude area composite required for entry to the MOS, the minimum acceptable score on that composite, whether or not the MOS was open to non-HSDG recruits, and whether or not it was open to females. This approach yielded 66 job clusters with 1986 accessions ranging from 30 to in excess of 17,000.

Our MOS sample included 121 MOS with total 1989 NPS accessions requirements of approximately 106000. The MOS not in the sample were excluded most often because their cluster had too few accessions over the 1982-1984 time period to allow for stable estimates of attrition probabilities. In addition, a number of MOS fell out of the analysis simply because we were unable to track changes in MOS designations across years.

VI. The Selection and Assignment Model

General Approach. Our objective is to develop a model that reflects the costs and payoffs of the first enlistment term based on current enlistment practices. Because current practices do not reflect optimal classification and assignment, our results cannot be characterized as "optimal". However, they do reflect cost-effective manpower utilization given existing recruiting practice. In the context of our model we specifically address the issue of how much recruit quality is cost-effective, and how does the cost effective mix of recruit characteristics vary as the number of recruits needed to meet total Army manpower requirements changes.

Figure 1
Network Formulation of Recruit Selection and Allocation Model
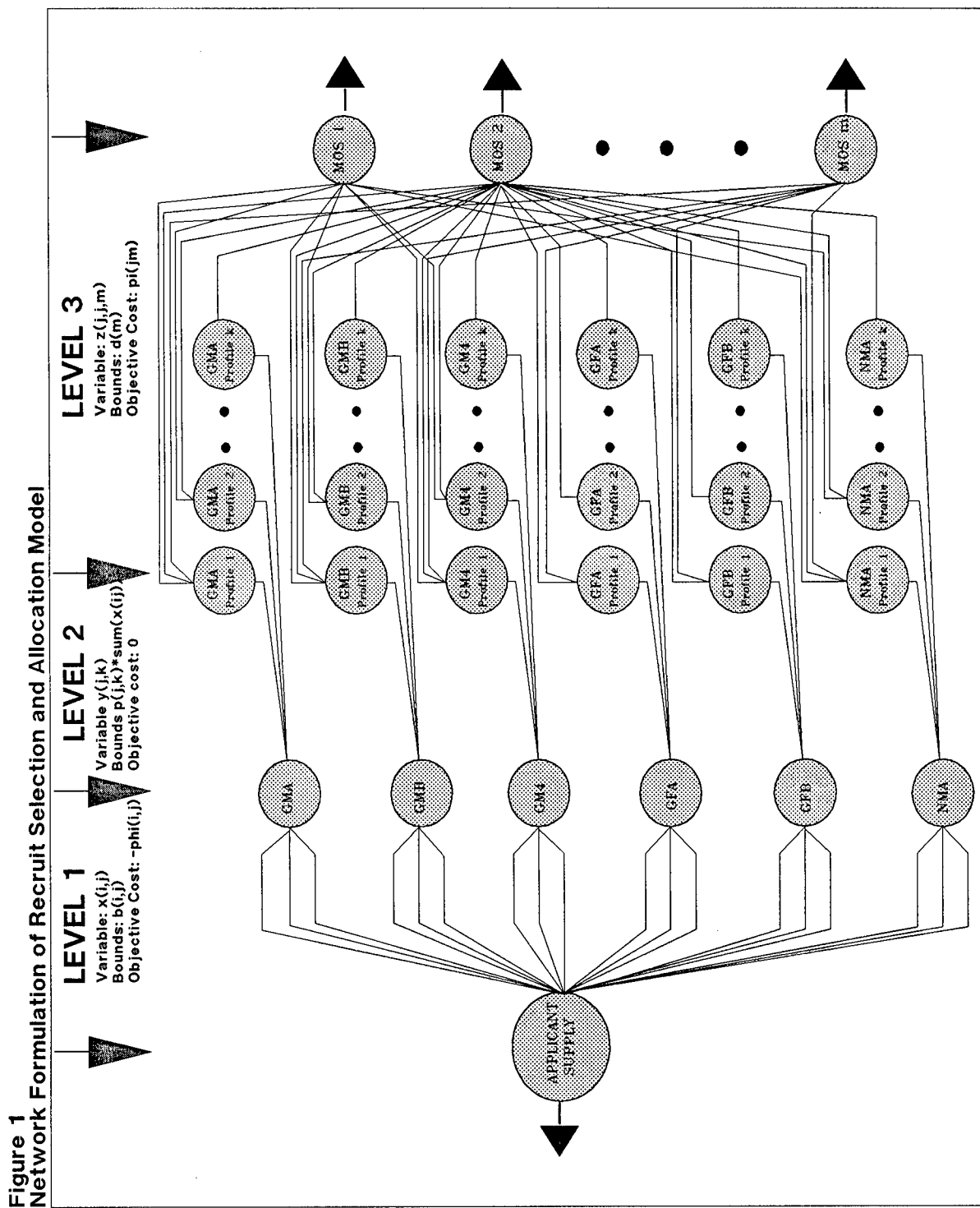
USAREC's objective can be conceptualized as maximizing π, net performance value, subject to limitations imposed by the force structure, limitations in population and ϕ, recruiting costs.

Figure 1 provides a graphic depiction of the allocation problem in the form of a network. The problem actually solved must be specified as a linear program (LP), rather than a network, but the network formulation provides a convenient way to conceptualize the problem's structure. The LP specification, using the same notation as that used in Figure 1, follows the discussion of the figure.

The network can be divided into three levels, each representing a different aspect of the selection and allocation problem. The arcs in this network represent variables in the optimization; constraints are represented by both nodes and bounds on the flows across specific arcs. For ease of exposition the flows at each level are denoted by different variables.

Recruiting costs are modelled in Level 1 of the network. The flows at this level, denoted x_{ij}, represent the numbers of accessions in mission box j recruited at average cost ϕ_{ij}, where i indexes segments of a piecewise approximation to the cost curve for mission box j. The objective function coefficient on the flow x_{ij} is $-\phi_{ij}$. Each arc in this level is constrained by an upper bound b_{ij}, representing the maximum flow that can be obtained at average cost i. The arcs are ordered so that for $i>i'$, $c_{ij}>c_{i'j}$, to represent the fact that average costs increase with the number of accessions in the mission box. The sum of the bounds ($\Sigma_i\, b_{ij}$) is the total available pool of potential applicants in the mission box. At optimality, the sum of all flows into a given mission box will be the cost-effective number of recruits for that mission box.

Level 2 models the distribution of recruits within mission boxes among the eligibility profiles. Here the flows, denoted y_{jk}, are the numbers of accessions in each mission box, j, that can be expected to fall into each eligibility profile. The constraints needed to model these flows are non-network constraints of the form:

$$p_{jk} \sum_i x_{ij} - y_{jk} = 0 \qquad (8)$$

where p_{jk} is the proportion of recruits in mission box j with eligibility profile k ($p_{jk}\geq 0$ and $\Sigma_k\, p_{jk}=1$ for all j,k.). The effect of this constraint is to force the optimization to replicate the actual recruiting system, which must pay the costs of recruiting and processing applicants who are discovered to be ineligible for available jobs late in the recruiting process.

In Level 3, the flows z_{jkm} represent the actual assignment of individuals from mission box j and eligibility profile k to job m. The "j" subscript is retained in this level to allow for the fact that, in the current specification, the performance payoffs in each job are assumed to be a function of mission box. These payoffs, denoted π_j are calculated using equation (6), and appear as the objective function coefficients on the variables z_{ijk}. The only purpose of the eligibility profiles in this specification is to enforce the minimum AA scores for each MOS.

The constraints on flows in this level of the network enforce MOS accession demands (d_m) by requiring $\Sigma_{jk}\, z_{jkm}=d_m$ for all MOS. Note that the sum of flows from a given eligibility profile to all MOS is not constrained to equal the total flow into that eligibility profile. The difference between these two quantities represents the number of recruits with that profile who are rejected because they cannot meet the minimum standard for any available job.

The structure described above can be mathematically represented as the following maximization problem:

Notation in the following equations has the same interpretation as in the preceding discussion, except that the distinction between levels 2 and 3 is eliminated -- flows from a mission box j through eligibility profile k to MOS m are represented as y_{jkm}.

$$\text{Maximize} \quad \sum_j \sum_m \pi_{jm} \sum_k y_{jkm} - \sum_i \sum_j \phi_{ij} x_{ij}$$

$$\text{Subject to:} \quad x_{ij} \leq b_{ij} \qquad \forall \ i,j$$

$$p_{jk} \sum_i x_{ij} - \sum_m y_{jkm} = 0 \qquad \forall \ j,k$$

$$\sum_j \sum_k y_{jkm} = d_m \qquad \forall \ m$$

$$x_{ij}, \ y_{ikm} \geq 0 \qquad \forall \ i,j,k,m$$

Computational Considerations: As specified, this problem has a maximum of $i*j+j*k*m$ variables, and $i*j+j*k+m$ constraints. Actual size is considerably smaller than this, however, because variables with $p_{ij}=0$ as well as those representing infeasible matches of MOS and eligibility profile can be eliminated from the model. For this analysis, the ranges of the indices were as follows: $i=1$-10, $j=1$-6, $k=1$-147, and $m=1$-66, resulting in a maximum of 58,278 variables, and 1014 constraints. The elimination of superfluous variables reduced this to 13,120 variables, and 473 constraints. The problem was solved using the SAS OR procedure PROC LP. SAS imposes limits on maximum record length that prevent creation of a matrix this large. To evade this constraint, we formulated the dual of the above problem, and used PROC LP to solve that. This strategy allowed us to obtain solutions to the problem, but increased computation time -- solution times ranged from 800-1200 seconds of CPU time.

Table 3
Estimated Cost Effective Selection Policy under Two Accession Missions

Mission Box	1988 Actual Mission		1990 Anticipated Mission	
	Number	Percent	Number	Percent
GFA	6,923	6.2	6,068	6.5
GFB	2,794	2.5	920	1.0
GMA	65,582	59.1	60,436	64.7
GMB	23,167	20.9	17,745	19.0
GM4	5,367	4.8	3,677	3.9
NMA	7,167	6.5	4,552	4.9
TOTAL	111,000	100.0	93,400	100.0
I-IIIA Percent		71.8		76.1

VII. Preliminary Results

The first application of the model detailed in this paper was a comparative analysis of cost effective selection policies under two different accession missions -- the 1988 total mission of 111,000, and an anticipated 1990 mission of 93,400. We are still in the process of completing this analysis, focusing particularly on evaluating the sensitivity of the results to variations in recruiting cost assumptions, but the preliminary findings are presented here as an illustration of the output of the model. Table 3 shows these results.

These results suggest that, as accession missions get smaller, it is cost-effective to increase the average quality content of the recruit pool. Note, however, that the optimization recommends decreases in the absolute mission for all categories of recruits. The recommended number of GMB recruits is unexpectedly low -- well under the both actual 1988 and currently missioned 1990 numbers for this group, while the number of GMA is higher than expected.

These preliminary results provide a starting point for continued analysis and refinement of the optimization model, as well as further development of the recruiting cost estimates. We are currently undertaking a series of simulations intended to answer at least some of the following questions:

(1) How robust are the optimization results under changes in estimated recruiting cost elasticities?

(2) What is the effect of using the standard assumption of constant marginal costs for non-quality recruits?

(3) How do the optimization results change if the objective of the problem is redefined to minimize costs subject to maintaining a fixed level of performance, rather maximize net performance value?

(4) What is the sensitivity of the results to changes in minimum job entry standards?

ENDNOTES

1. Jane M. Arabian and Jeanne K. Mason, "Relationship of SQT Scores to Project A Measures," Proceedings of the Military Testing Association Conference, June 1987. p. 351.

2. A partial solution to this difficulty with SQT was first proposed in a series of papers known as the Rand model (See Armor(1981); Armor, Fernandez, Bers, and Schwarzbach (1982); and Fernandez and Garfinkel(1985)).

3. eg. artillery gunners with higher scores tend to acquire targets more quickly; tank gunners tend to score more hits.

4. See David Armor, Richard Fernandez, Kathy Bers, and Donna Schwarzbach, Recruit Aptitudes and Army Job Performance: Setting Enlistment Standards for Infantrymen, The Rand Corporation, R-2874-MRAL, September 1982; Richard L. Fernandez and Jeffrey B. Garfinkle, Setting Enlistment Standards and Matching Recruits to Jobs Using Job Performance Criteria, The Rand Corporation, R-3067-MIL, January 1985.

5. Frances C. Grafton and David K. Horne, An Investigation of Alternatives for Setting Second-to-Third Tour Reenlistment Standards, The U.S. Army Research Institute, Technical Report 690, August 1985.

6. D.H. Mclaughlin and P.G. Rossmeissl, L.L. Wise, D.A. Brant, and M.M. Wang, Validation of Current and Alternative ASVAB area Composites, based on training and SQT information of FY82 and FY82 enlisted Accessions. Technical Report 651. Alexandria VA: U.S. Army Research Institute.

7. C.H. Campbell, R.C. Campbell, M.G. Rumsey, and D.C. Edwards, Development and field test of job-relevant knowledge tests for selected MOS-specific criterion measures, Technical Report 717. Alexandria, VA.: Army Research Institute

8. Joseph Zeidner and Cecil Johnson, The Economic Benefits of Predicting Job Performance, IDA paper P-2241, Alexandria VA: Institute for Defense Analyses.

9. Frank L. Schmidt and John Hunter , "Individual Differences in Productivity: An Empirical Test of Estimates Derived From Studies of Selection Procedure Utility", Journal of Applied Psychology, Vol 68, no. 3, 1983, 407-414.

10. Excluded from the sample were those with prior military service.

11. See Military Compensation Background Papers, 3rd Edition (Washington, DC: Department of Defense, Office of the Secretary of Defense, June, 1987)

12. While there is greater variation in the time in grade for E-3s before promotion to E-4, much of this variation appears to be related to differences in turnover rates for different MOS rather than differences in productivity.

13. This assumes that the effects of expenditures on other recruiting resources is similar to that of wages. This seems the most reasonable approach since dramatic differences in the return to alternative recruiting resources would suggest efficiency gains from increasing expenditures on alternative resources and reducing military salaries for the lower ranks. This also the estimated slope used by AMCOS.

14. Since military pay increases can be influenced by military budget requests, it is not completely endogenous. Nevertheless, Congress exercises a great degree of autonomy in establishing pay raises for federal employees.

15. Processing costs include the costs of recruiter time and recruiting resources necessary to complete and check applications and the costs of screening applicants. This includes the costs of administering both the ASVAB and a physical examination.

16. To avoid this cost, the Army temporarily lowered minimum standards for some combat MOS to insure that otherwise eligible non-quality applicants could be placed and recruiter time would not be wasted and transferring the cost from recruiting to reduced performance

17. Paul F. Hogan, Donald E. Rose, Jr., Richard W. Hunter, Lee S. Mairs, and Jeffery A. Zuckerberg, AMCOS Information Book, Technical Report, (Arlington, VA: Systems Research and Applications Corp.), March 30, 1987.

The Sergeant Major Study Continued: Objective Health Risk Assessment
By Laboratory Blood Tests

Richard C. Keniston, MAJ, MC*
Gerald Cross, LTC, MC**
John I. Enriquez, Sr.*
Ismael Delgado, SP5, MC*
Fred Duncan#
Department of Clinical Investigation*,
Resident Treatment Facility**, and
Health Fitness Center#,
William Beaumont Army Medical Center
El Paso, Texas 79920-5001

As we discussed in our previous article(1), the Army is concerned about preserving the health of its soldiers, their dependents, and retirees. For this purpose, periodic physical examinations, and a health risk screening program have become part of the Army's health care system. This health risk screening program, the Health Risk Appraisal (HRA), relies heavily on subjective information (the results of a questionnaire). Objective measurements include height, weight, blood pressure, and two blood chemistries, fasting glucose and total cholesterol. In our experience, the HRA as currently configured only identifies about one-half of the subjects truly at risk, as determined by follow-up studies of who has a heart attack, gets into a motor vehicle accident, or develops cancer (1,2). Because the HRA relies on tests with limited sensitivity and specificity, such as total cholesterol, and also relies on self-reporting, which can be inaccurate, it fails in its mission to perserve the health of the fighting force.

We were interested in seeing if we could improve upon the reliability of the HRA. We reasoned that people at greater risk to develop serious health problems, such as cigarette smokers, alcoholics, or drug abusers, ought to be different metabolically from healthy nonsmokers, non-heavy drinkers (the lowest risk group)(3). In our first article (1), we reported

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

that this was indeed the case. We could classify approximately 90% of the sergeant major candidates correctly as to cigarette smoking status on the basis of blood tests alone. Heavy drinkers of alcohol could likewise be reliably identified on the basis of blood tests. Similarly, those who were destined to develop serious health problems within the study period were shown to be metabolically closer to cigarette smokers and heavy drinkers than were those not so destined. We wondered if a more universal health risk status evaluation system could be developed, a system that would be completely objective, and not depend upon any input from the person being evaluated, other than his cooperation in obtaining the blood samples.

So, we decided to look at other groups of people, who would likely not be in even as good health as the smoking or heavy-drinking sergeant major candidates. One such group were the drug/alcohol rehabilitation patients from the Resident Treatment Facility. Another such group were hospital inpatients; we obtained cooperation from the General Surgery Service, and included these patients as being representative of people in relatively poor health. We thus obtained a broad spectrum of subjects, from those in general good health through cigarette smokers and heavy drinkers, drug/alcohol rehabilitation patients, outpatients, inpatients, seriously ill and very seriously ill inpatients, all the way up to patients destined to die during the study period. The goal was to measure substances in the blood of all these people, and then derive a mathematical model for ranking these people by health risk status, based solely on the results of the blood tests.

METHODS

Study Population: Most of the 702 subjects in the main study group came from three sources: (1) Sergeant Major Candidates (SGM) (n=354); (2) Resident Treatment Facility Patients (RTF) (n=102); and (3) Hospital Inpatients (n=224) (85% from General Surgery Service). The other 22 subjects were healthy laboratory controls and outpatients. All subjects who were having their blood drawn for other than clearly indicated medical reasons were asked to give informed consent, under research protocols approved by the WBAMC Clinical Investigation and Human Use Committees.

Subjects used for validation of the derived health risk status

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

equation came primarily from four sources: (1) Cardiology (n=88); (2) Trauma Unit (n=27); (3) Neonatal ICU (n=32); and (4) Renal Dialysis Unit (n=10).

Subjective Data: Healthy subjects were asked to fill out a questionnaire relating to their smoking and drinking habits, use of drugs, vitamins, and other medications, acute or chronic health problems, exercise, diet, and fasting status prior to the blood drawing. Hospital patients were interviewed when possible, and their charts were reviewed for the type of information obtained by questionnaire from the healthy subjects.

Objective Data: Blood was drawn from the healthy subjects after an overnight fast. The sergeant major candidates had blood drawn twice, at a 5-month interval. The hospital patients had blood drawn for screening or diagnostic purposes as part of their medical therapy. The following tests were run on the collected blood samples: (1) Serum chemistry screen (SMAC-20) (see reference 1 for a list of the tests); (2) Complete blood count (CBC); and (3) vitamin B6 (plasma pyridoxal 5'-phosphate, PLP).

Derivation of Health Risk Status Prognostic Equation: The 702 main study group subjects were classified by **Actual Health Risk Status**, as outlined in Table 1. Stepwise multiple regression analysis was then used to solve for the dependent variable, **Actual Health Risk Status Score**, employing the 20 serum chemistry tests, 8 complete blood count tests, and plasma PLP as the independent variables. The computer program employed was Statview 512+. The prognostic equation was used to calculate **Predicted Health Risk Status Scores** for the subjects, and the predicted scores were compared to the actual scores.

Validation of the Derived Prognostic Equation: We looked at four groups of hospitalized subjects (cardiology patients, trauma patients, premature infants, and renal dialysis patients) who were not used in the derivation of the prognostic equation, and used their blood test results to calculate their **Predicted Health Risk Status Scores**. These predicted scores were compared to the actual scores.

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

RESULTS

Sex, Age, Cigarette Smoking, And Alcohol Status of Subjects:
The sex percentage of each **Health Risk Status Score** group is indicated
in Table 2, as is the age distribution data. The groups varied widely both
in percentage of females, and in median age. The cigarette smoking and
alocohol status of the subjects is indicated in Table 3; again, as would be
expected when these variables were used as grouping criteria, the **Health
Risk Status Score** groups varied widely in percentage of subjects who
smoked and drank to excess.

**Ability of Predicted Health Risk Status Score to Correctly
Classify the Main Study Subjects as to Actual Health Risk
Status Score:** Table 4 and Figure 1 show the relationship between
Predicted Health Risk Status Score and **Actual Health Risk
Status Score.** The correlation coefficient (r) = + 0.8888, F = 2633.7,
and p < .00001, indicating a highly significant relationship between the
Predicted and the **Actual Health Risk Status Scores.** Using the
indicated **Predicted HRSS** cutoff scores (Table 5), 5 of 9 (55.56%) of
subjects were classified precisely to the correct **Actual HRSS** category,
and another 31.47% were classified to within one category of their
Actual HRSS category.

There were a number of significant outliers (Figure 1 and Table 4).
The one subject in category 3 who had a **Predicted HRSS** of nearly 6 was
a drug/alcohol rehabilitation patient with acute alcoholic hepatitis. In
category 4, two of those with **Predicted HRSS** scores greater than 6
ultimately died at greater than six months. In that same category
(category 4), many of those with **Predicted HRSS** scores less than 2.5
were in for elective surgeries, such as hernia repair. The six subjects in
categories 6 and 8 with **Predicted HRSS** scores less than 4 were all
supposedly healthy at the time their **Predicted HRSS** scores were
determined; nevertheless, their screening **Predicted HRSS** scores
averaged much higher than the **Predicted HRSS** scores of those who did
not subsequently develop serious, life-threatening health problems
(2.225+-0.28 vs 1.74+-0.48, t = 3.794, p < .001). The three subjects in
category 10 with **Predicted HRSS** scores less than 5 all died within 72
hours of admission, and all died of acute problems.

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Ability of Predicted Health Risk Status Score to Correctly Classify Other Groups of Subjects as to Actual Health Risk Status Score: Figures 2, 3, and 4 show the relationship between **Predicted Health Risk Status Score** and **Actual Health Risk Status Score** in three different groups of hospital patients not used in the derivation of the prognostic classification equation. These three groups are (1) Cardiovascular Disease patients (Figure 2), (2) Trauma patients (Figure 3), and (3) Premature Infants from the NICU (Figure 4). In each group, there is a highly significant correlaion between **Predicted Health Risk Status Score** and **Actual Health Risk Status Score.** For the Cardiovascular Disease patients, $F = 118.676$, and $p < .0001$. For the Trauma patients, $F = 65.48$, and $p < .0001$. For the Premature Infants, $F = 26.332$, and $p < .0001$. For 10 Renal Dialysis patients, one of whom died, $F = 30.45$, and $p < .0001$. Thus, even in groups of patients not used in its derivation, the prognostic classification equation was useful in estimating **Actual Health Risk Status.**

The Premature Infants were the only group in which any subject with a **Predicted Health Risk Status Score** of greater than 9 survived. In the Premature Infants, the cutoff **Predicted HRSS** best separating the survivors from the deceased was 9, instead of 7, as it was for the other groups. The category 10 infant with the lowest **Predicted HRSS** died of multiple congenital anomalies, rather than just respiratory and metabolic problems (Figure 4).

DISCUSSION

We have derived an equation for estimating health risk status in a totally objective fashion. This estimate relies only on the results of blood tests, not on any subjective input from either the subject or the Army or medical personnel caring for him. This estimate takes into account derangements in the blood tests induced by cigarette smoking, alcohol and drug abuse, use of medications, and medical and surgical therapies. The estimate is independent of sex, age, and medical diagnosis, as these factors were not used in its derivation. These equation is useful for predicting health risk status in groups as diverse as healthy middle-aged men, elderly cardiovascular disease patients, surgical patients, young adult trauma patients, and premature infants.

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

The most important single blood test was serum albumin, which explained 2/3 of the total variance, and 80% of the explainable variance. The next most important test was total carbon dioxide (mainly bicarbonate), which is slightly reduced in cigarette smokers and markedly reduced in seriously ill patients. The third most important test was log(10) of plasma PLP (vitamin B6), which falls off in direct proportion to the severity of illness, and which is the second most important factor, after white blood cell count, for discriminating between cigarette smokers and nonsmokers. Some tests, such as phosphate and gamma glutamyltransferase, were related to alcohol status, with alcoholics having higher values. What was significant was the absence of total cholesterol as an independent discriminator of health status; total cholesterol averaged lowest in those destined to die (category 10), and a markedly elevated total cholesterol (> 280 mg/dL) was associated with only a doubling of six-month mortality, relative to the reference range of from 120 to 280 mg/dL.

When combined with our previously reported profiling instrument, the **Health Risk Assessment Profile** (HRAP)(1), the **Predicted Health Risk Status Score** has the potential to revolutionize health risk assessment. Apparently healthy people destined to soon have serious health problems diagnosed usually have elevated values for the **Predicted HRSS** (See the six subjects illustrated in Figure 1). Smokers, long known to be at approximately double the risk of nonsmokers (3), are closer in average **Predicted HRSS** to the hospital inpatients than are the nonsmokers. Serial determinations of **Predicted HRSS** may also prove useful - a subject whose score increased over time may be developing an occult health problem (We have three cases where this occurred - one case of malignant melanoma, one case of lung carcinoma., and another case of intraductal carcinoma-in-situ of the breast). The **Predicted Health Risk Status Score,** based as it is on many different blood tests, reflects changes in any one or more of these tests. Thus, subtle changes that would not normally picked up by the health risk screeners or screening physician might nevertheless result in a notable increase in the **Predicted HRSS.** All test results might even be well within the reference ranges, but the total pattern would be abnormal.

The use of the **Predicted HRSS** is not limited to health risk screening, or long-term individual patient followup. It could be useful for acute patient care. For instance, one would expect a surgical patient's

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Predicted HRSS to deteriorate somewhat immediately after surgery. However, if the patient is improving after surgery, his **Predicted HRSS** should fall again in the post-op recovery period. If the **Predicted HRSS** did not improve or even continued to deteriorate (rise), an occult health problem could be present. We have seen many autopsies where a problem, diagnosable by blood test results, was missed because the none of the individual test results were that abnormal.

The **Predicted HRSS** could also be used to monitor drug, medical, or dietary therapies. If the the overall net effect of the therapy were beneficial, the **Predicted HRSS** should fall, or at least not rise. Standard low-salt and lipid-lowering diets may have their benefits outweighed by inducing vitamin and mineral malnutrition (4), as we saw with vitamin B6 and calcium levels in the Sergeant Major candidates (1). These adverse effects on nutritional status may explain why none of the lipid-lowering studies has reported a net decrease in all-cause mortality (5,6).

The major drawback to widespread use of the **HRAP/Predicted HRSS** instrument is its cost - it requires more tests to be run, and reagents and technician time cost money. On the other hand, the Army spends millions of dollars maintaining its hardware in an effective fashion. So, wouldn't it be wise to make certain that the operators of that expensive equipment and their leaders also were in optimal condition? Our experience with the Health Risk Appraisal (HRA) is that it does not reliably meet the goals that were set for it - it identifies only about one-half of the people who subsequently develop serious health problems as being at increased risk. We are confident that the **HRAP/Predicted HRSS** instrument could do much better than that.

REFERENCES

1. Keniston RC, Weir MR, Enriquez JI Sr, Duncan F. The sergeant major study: Health risk assessment by clinical laboratory parameters. 1988 Army Science Conference Proceedings. Vol. II:47-61.

2. Keniston RC. An isolated total cholesterol level is inadequate. JAMA 1989; 261:3241-3242.

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

3. Klatsky AL, Friedman GD, Siegelaub AB. Alcohol and mortality, a ten-year Kaiser-Permanente experience. Ann Int Med 1981; 95:139-145.

4. Engstrom AM, Tobelman RC: Nutritional consequences of reducing sodium intake. Ann Int Med 1983; 98:870-872.

5. ---. The Lipid Research Clinics coronary primary prevention trials results. I. Reduction in incidence of coronary heart disease. JAMA 1984; 251:351-364.

6. Frick MH, Elo E, et al. Helsinki heart study: Primary prevention trial with gemfibrozil in middle-aged men with dyslipidemia. N Engl J Med 1987; 317:1237-1245.

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Table 1. HEALTH RISK STATUS SCORE CATEGORIES

Category	Subjects Included in Category	Number in Main Study
1	Healthy nonsmokers, non-heavy drinkers	212
2	Healthy cigarette smokers, non-heavy drinkers or nonsmokers, heavy drinkers (> 30 drinks of alcohol/month, but no alcohol-related social or medical problems)	145
3	Cigarette smokers, heavy drinkers or drug/alcohol rehabilitation patients or medical center out-patients or pregnancy	121
4	Hospital inpatients, not otherwise categorized	119
5	Renal hemodialysis patients, for instance	0
6	Seriously ill hospital patients	37
7	Acute trauma or acute myocardial infarction or burn patients, for instance	0
8	Very seriously ill hospital patients, lived six months	14
9	Premature infants in neonatal intensive care unit, for instance	0
10	Died within six months	54

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Table 2. SEX AND AGE OF SUBJECTS

Actual Health Risk Status	Sex (Percent)		Age (Years)			
	Male	Female	Mean	1 S.D.	Median	Range
1	98.11	1.89	38.9	3.87	39	6-55
2	97.93	2.07	38.5	3.29	38	32-50
3	79.34	20.66	34.6	10.0	35	7-70
4	50.42	49.58	50.3	16.4	56	7-84
6	70.27	29.73	53.6	15.9	56	16-79
8	78.57	21.43	49.6	13.7	48	23-68
10	66.67	33.33	57.8	17.6	64.5	0-89
Combined	82.48	17.52	42.4	8.82	42	0-89

Table 3. CIGARETTE SMOKING AND ALCOHOL STATUS OF SUBJECTS (Percent)

Actual Health Risk Status	Cigarette Smoking Status			Alcohol Status		
	Never Smoked	Exsmoker	Current Smoker	Non-Drinker	Moderate Drinker	Heavy Drinker
1	58.02	41.98	0.0	26.42	73.58	0.0
2	4.14	4.83	91.03	15.17	75.86	8.97
3	11.57	7.43	80.99	4.96	5.79	89.26
4	44.53	24.37	31.09	39.50	46.22	14.28
6	24.32	18.92	56.76	18.92	51.35	29.73
8	21.43	0.0	78.57	35.71	21.43	42.86
10	33.33	27.78	38.89	40.74	25.93	33.33
Combined	32.70	21.02	46.28	23.58	52.41	24.02

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Table 4. DISTRIBUTION OF SUBJECTS BY PREDICTED AND ACTUAL HEALTH RISK STATUS SCORE CATEGORIES

Predicted Health Risk Status Score Category	Actual Health Risk Status Score Category						
	1	2	3	4	6	8	10
< 1.500	105	19	10	2	0	0	0
1.500-1.999	76	46	14	7	1*	0	0
2.000-2.399	23	48	28	16	0	1*	0
2.400-2.999	8	28	47	21	2*	1*	0
3.000-4.499	0	4	21	54	6*	0	0
4.500-5.999	0	0	1	15	11	0	4
6.000-6.999	0	0	0	3	7	9	8
7.000 or higher	0	0	0	1#	10#	3#	42

* Includes subjects who had these predicted health risk status scores at the time of health risk screening and who subsequently developed serious health problems (3 heart attacks, 3 alcohol-related motor vehicle accidents).
# Includes subjects who died at greater than six months.

Chi-square = 1235.952, p < .0001, 174 df

Predicted Health Risk Status Score = 10.998 - .081*albumin (g/L) - .056*total CO2(mEq/L) - 1.312*log(10) of PLP (vitamin B6)(nmol/L) + .269*phosphate (mg/dL) - .358*red blood cells (10^6) + .407*potassium (mEq/L) + .002*gamma glutamyltransferase (U/L) - .325*calcium (mg/dL) + .001*fasting triglycerides (mg/dL) + .001*lactic dehydrogenase (U/L) + .003*PLP (vitamin B6)(nmol/L) + .001*platelets (10^3) + .002*fasting glucose (mg/dL) - .067*uric acid (mg/dL) + .016*white blood cells (10^3)

Percent of subjects correctly classified by the above equation to precise Actual Health Risk Status Score = 55.56% (390 of 702).
Percent of subjects correctly classified by the above equation to within one category of correct category = 87.03% (611 of 702).

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

TABLE 5. ABILITY TO DISCRIMINATE BETWEEN VARIOUS HEALTH RISK STATUS SCORE (HRSS) CATEGORIES

1st HRSS Category	2nd HRSS Category	HRSS Cutoff	Predicted HRSS Means	1 S.D.	Medians	Significance F-value	p
1	2	2.000	1.55, 2.04	0.43, 0.50	1.55, 2.05	9.556	<.0001
2	3	2.400	2.04, 2.49	0.50 0.66	2.05, 2.49	6.136	<.0001
3	4	3.000	2.49, 3.38	0.66, 1.19	2.49, 3.19	7.119	<.0001
4	6	4.500	3.38, 5.67	1.19, 1.86	3.19, 5.52	6.965	<.0001
6	8	6.000	5.67, 6.24	1.86, 2.39	5.52, 6.54	0.779	>.10
8	10	7.000	6.24, 8.79	2.39, 2.01	6.54, 8.73	3.551	<.001

1st HRSS Category	2nd HRSS Category	HRSS Cutoff	Sensi-tivity	Speci-ficity	PV of POS	PV of NEG	Effi-ciency
1	2	2.000	55.17	85.38	72.07	73.58	73.11
2	3	2.400	57.02	80.69	71.13	69.23	69.92
3	4	3.000	60.50	81.82	76.60	67.81	71.25
4	6	4.500	72.97	84.03	58.70	90.91	81.41
6	8	6.000	85.71	54.05	41.38	90.91	62.75
8	10	7.000	77.78	78.57	93.33	42.83	77.94
1	10	4.000	100.0	100.0	100.0	100.0	100.0
2	10	4.000	100.0	100.0	100.0	100.0	100.0
1	3	2.000	80.17	85.38	75.78	88.29	83.48
1	4	2.400	78.15	96.23	92.08	88.70	89.73
1	6,8	3.000	88.24	100.0	100.0	97.25	97.72
2	4	2.400	78.15	80.68	76.86	81.82	79.54
2	6,8	3.000	88.24	97.24	91.84	95.92	94.90

KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Figure 1. Correlation between Actual and Predicted Health Risk Status Scores in 702 main study group subjects

$$y = 1.008x + .165, \text{ R-squared: } .79$$
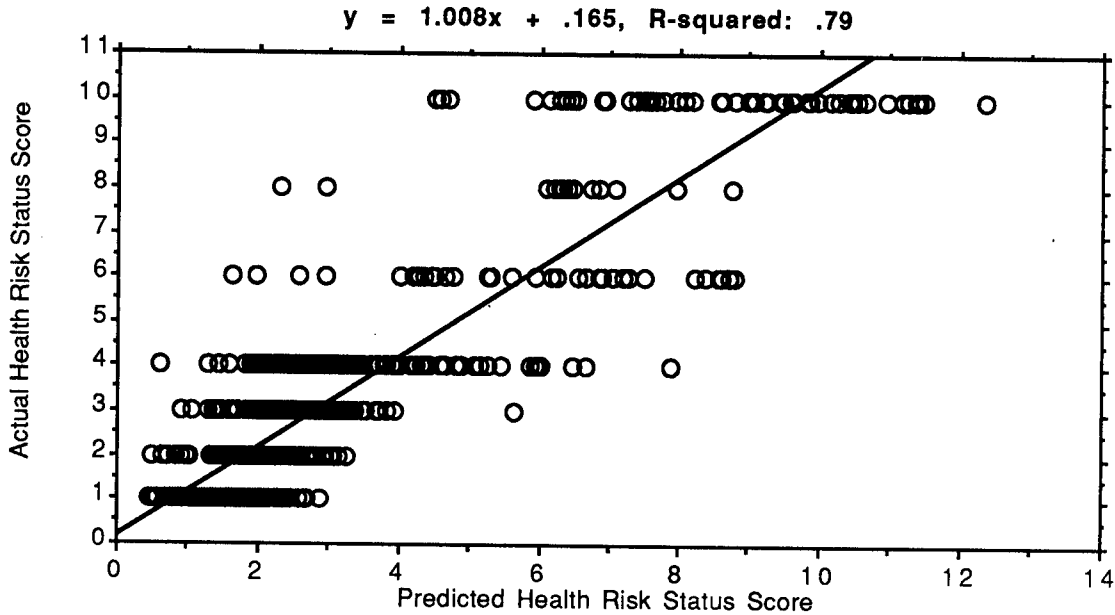
Figure 2. Correlation between Actual and Predicted Health Risk Status Scores in 88 cardiovascular disease patients not used in derivation of prognostic equation

$$y = .839x + 1.902, \text{ R-squared: } .577$$
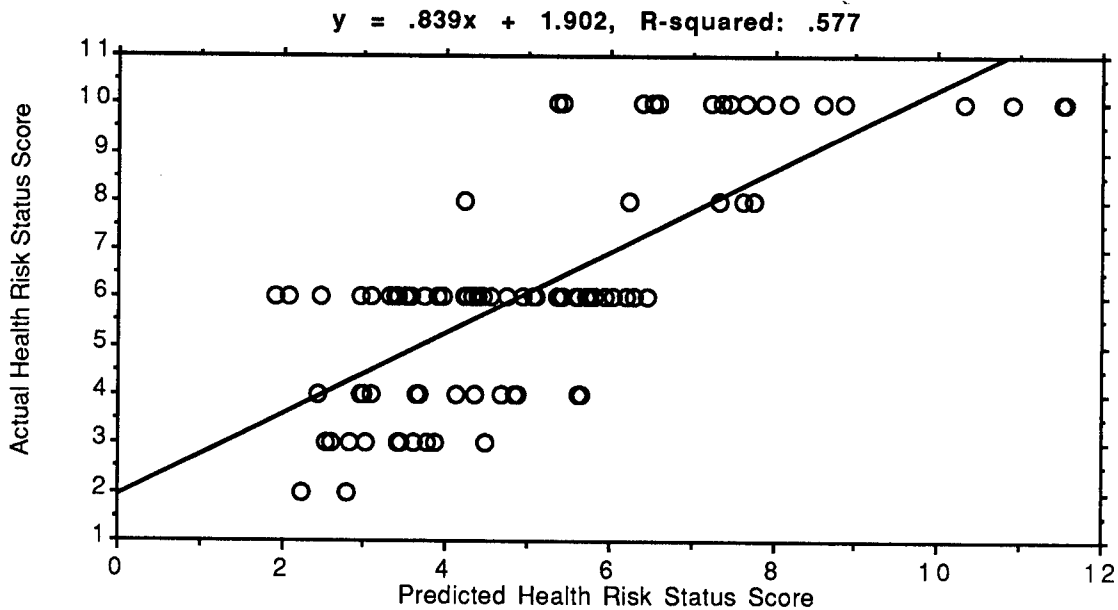
KENISTON, CROSS, ENRIQUEZ, DELGADO, & DUNCAN

Figure 3. Correlation between Actual and Predicted Health Risk Status Scores in 27 trauma patients not used in derivation of prognostic equation

$y = 1.09x + .161$, R-squared: .716
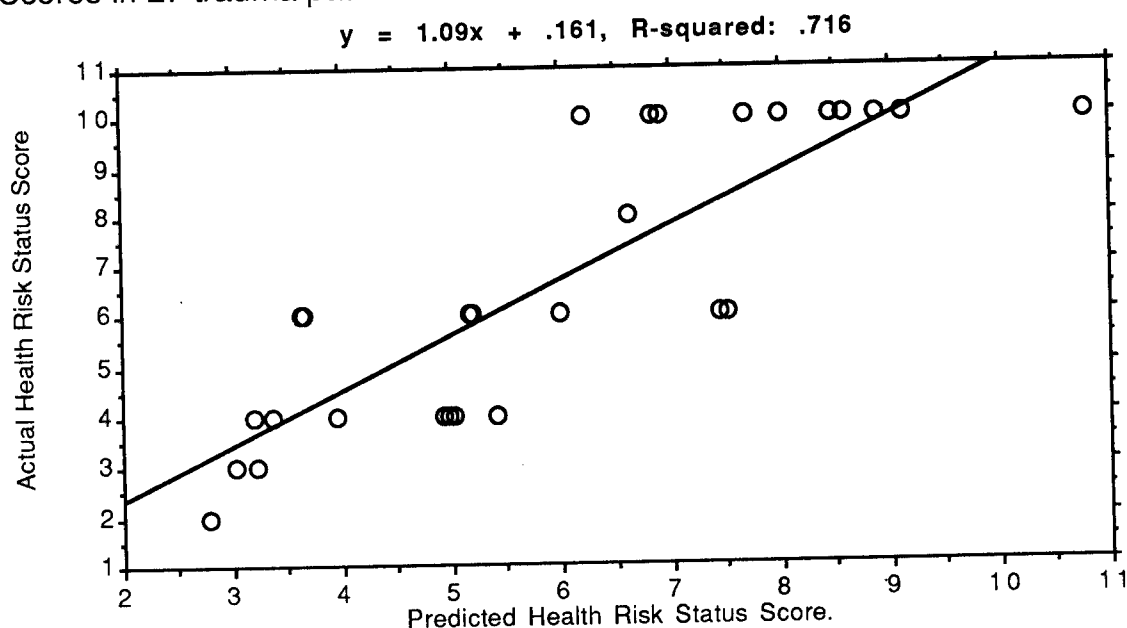
Figure 4. Correlation between Actual and Predicted Health Risk Status Scores in 32 premature infants not used in derivation of prognostic equation

$y = .634x + 3.105$, R-squared: .467
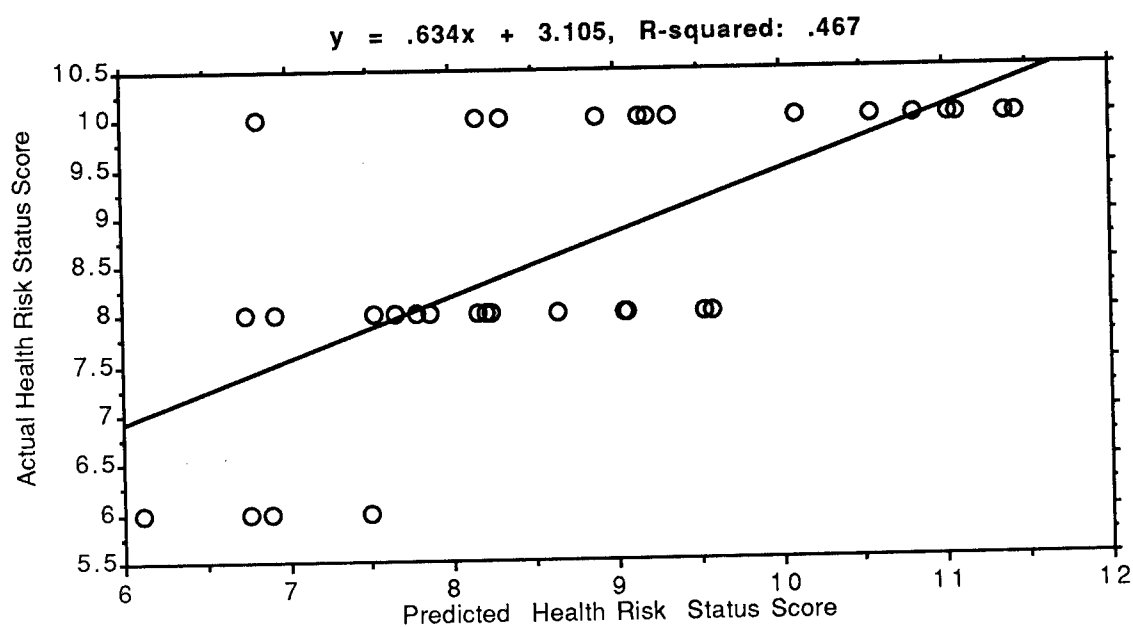
Intrinsic Chemical Markers for Thermal Processing
of Particulate Foods

*Hie-Joon Kim, Dr., and Irwin A. Taub, Dr.
Food Engineering Directorate
US Army Natick Research, Development and Engineering Center
Natick, MA 01760-5018

INTRODUCTION

Since Nicholas Appert invented canning in 1809, industrial thermal processing of foods has remained primarily a batch process. Currently, it is undergoing a revolutionary change through the introduction of aseptic processing, in which fluid foods of uniform consistency are sterilized while flowing through tubes maintained at a high temperature (between 130°C and 150°C) for a short time (HT/ST) and packaged in a sterilized container in a continuous process. Aseptic processing is widely used in drinks and semiliquid foods, but its use has not been approved for foods containing particulates. The Food and Drug Administration (FDA) requires demonstration of sterility at the center of a food particulate based on a definitive microbiological procedure and a conservative mathematical model.[1]

Recently, attempts have been made to use B. stearothermophilus, a thermophilic bacteria, as a bioindicator of aseptic processing of particulate foods.[2,3] Encouraging results have been reported. For example, Hinton et al. placed at the center of meat cubes small alginate beads inoculated with B. stearothermophilus and found no survivors after a short (0.87 min) processing at 131°C.[3] Based on this observation, they estimated the equivalent lethality for C. botulinum at the center of the particulate to be at least 25D (D is time required to destroy 90% of the microorganism), which represents a gross overprocessing compared to the required 12D for commercial sterility. The drawback of the microbiological procedure is that it is tedious and subject to many experimental uncertainties. Moreover, the extent of overprocessing is difficult to assess because, once the entire microbial population is destroyed, only a minimum lethality can be estimated.

The mathematical modeling approach, which involves calculating microbial destruction within the particulate as it flows through the holding tube, is subject to even more uncertainties. Various assumptions and estimates have to be made, because direct temperature measurement at

the center of a moving food particulate is extremely difficult. The major uncertainties relate to estimating the convective heat transfer coefficient at the liquid-particulate interface and the residence time distribution of the particulates in the holding tube.[4] Moreover, the conservative estimates used for these parameters, which are difficult to verify experimentally, often lead to unrealistically long processing time requirements.[5]

Monitoring chemical changes within the food, which involves compounds either indigenous to or added to the food, offers an alternative route for assessing the integrated time-temperature exposure at different locations within the food particulate. Published data in the literature are either scarce (for thermal processing in general) or nonexistent (for particulate foods). There are good reasons for this scarcity. A typical chemical reaction in foods is either too fast or too slow for the associated compound to be a useful indicator of sterility. For example, destruction (blanching) of heat-resistant enzymes, such as peroxidase and catalase, is complete in about 2 min at 100°C. On the other hand, destruction at high temperatures of such nutrients as ascorbic acid and thiamin is generally much slower than thermal death of bacterial spores. For example, the use of thiamin added to pea or beef puree in cans was suggested as an indicator of sterility for retorting at 122°C.[6] The z value, i.e., the degrees in F required to increase the destruction rate by 10, is 18 for C. botulinum, but it is 48 for thiamin at 122°C. Therefore, at aseptic processing temperatures, the destruction of thiamin will be too slow to be an adequate indicator of sterility. Incidentally, this higher retention of nutrients is one of the important advantages of aseptic processing.

To date, there is no information available in the literature that would point to a chemical reaction potentially useful for monitoring thermal processing of particulate foods at temperatures above 122°C. However, the desirability of such an approach has been recognized in the past. In 1975, Mulley et al. stated "while the whole area of chemical indicators promises to be a fertile ground for research and patent hunters, its practical application may be just beginning. The authors believe that the use of a chemical index in sterilization processing has the potential of effecting a revolutionary change in the food and pharmaceutical industries."[6] Despite this prophetic statement, not much progress has been made during the past 15 years.

Recently, we undertook a systematic investigation of chemical markers whose changes could effectively represent integrated time-temperature exposure at different locations within a food particulate. The desired features of the chemical markers were considered to be as follows:

(a) The chemical reactions should be intrinsic to the foods, because addition of chemicals to the foods would be tedious and would introduce experimental errors.

(b) The marker should be <u>thermally produced</u> instead of <u>destroyed</u>, because analytical results will be more reliable when the baseline concentration for unheated food is essentially zero.

(c) The marker should be easily <u>detectable</u>, because heating will generate a large number of compounds that when taken together with existing compounds in the foods make analysis more difficult.

(d) The analytical method for the marker should be <u>sensitive</u>, because the marker concentration at the center portion (less than 50 mg sample) of a food particulate (less than 1 inch in diameter) needs to be determined.

(e) The marker or its precursor should be <u>stable</u> toward oxidation or enzymatic reactions, because instability of the marker or the precursor could lead to inconsistent results in the marker concentration.

(f) The marker should be <u>common</u> to many foods to be suitable for broad applications.

(g) The marker should be a <u>time-temperature integrator</u> instead of a maximum temperature indicator, because microbiological lethality, F_o, corresponds to the accumulation of incremental lethalities at different time-temperature exposures.

(h) The reaction rate for the marker should follow <u>first-order kinetics</u> so that the marker reaction could be related to the thermal death of microorganisms, which is a first-order process.

(i) The marker concentration should show a <u>correlation with F_o</u> so that one could in principle verify F_o from the marker concentration.

In this paper we report the detection and isolation of two as yet unreported compounds that, based on the above-mentioned considerations, appear to be promising time-temperature integrators at aseptic processing temperatures.

DETECTION

Many analytical methods are available for determination of food components. Gas or liquid chromatographic techniques are most commonly used. Various chromatographic techniques with different detection methods are useful for detecting a different subset of compounds in foods. Since organic acids and carbohydrates represent an important class of compounds commonly present in most foods, we decided to investigate thermally induced changes in the profile of these compounds. Anion exclusion chromatography (AEC) is a useful high-performance liquid chromatographic (HPLC) technique for determination of organic acids and other carbohydrates such as ascorbic acid and various sugars. Since we were searching for markers without any knowledge of their properties, a general, instead of a specific, detection scheme was desired. Therefore, a photodiode array (PDA) detector, which is sensitive in the UV region and has scanning capability, was used. With it, the UV absorption spectra of compounds eluting from the chromatographic column could be

obtained every fraction of a second, stored in a computer, and manipulated for display as a three-dimensional (3-D) representation, a contour map, or a spectrum at a specific time. The 3-D nature of this anion exclusion chromatography-photodiode array (AEC-PDA) system proved extremely useful for discerning new compounds being produced as a result of heating.

Figure 1 shows the contour diagram for the spectrochromatogram (i.e., the absorbance vs. wavelength as a function of retention time) of unheated and heated broccoli. Wescan anion exclusion column (sulfonated polystyrene/divinylbenzene, 7.8x100 mm) and Waters 990 photodiode array detector were used. The eluant was 5 mM sulfuric acid solution and the flow rate was 1 ml/min. Comparison of the contour diagram for unheated (control) and heated broccoli clearly shows production of a new compound (M-1: retention time, 4 min; UV maximum, 300 nm) not present in unheated broccoli. The thermal reaction is intrinsic to the food, i.e., the new compound is produced without adding any compounds prior to processing. It is easily detectable because the compound has a relatively long retention time and a characteristic absorption maximum at 300 nm. No similar absorption peak is present in the wavelength-time domain of the control broccoli. The same peak was observed upon heating other vegetables (e.g., potato, green bean, pea, carrot) and fruits (e.g., apple, orange). The compound was also observed from heated meats (e.g., beef, chicken), even though an interfering peak with slightly shorter retention time and lower peak wavelength presented a potential problem in meats. Fortunately, another compound was discerned in heated meats (M-2: retention time, 6 min; UV maximum, 285 nm). These compounds were observed reproducibly. Unlike destruction of ascorbic acid, the thermal production of these compounds was not observed to be influenced by oxygen. They were also stable, which facilitates analysis and subsequent purification. We further investigated the usefulness as time-temperature integrators of M-1 for vegetables and of M-2 for meats .

CORRELATION WITH F_o

F_o is the time in minutes at 121.1^oC required to achieve "commercial sterility" by reducing the population of a tester spore-forming microorganism a specified factor, such as 10^6 for <u>PA3679</u> or 10^{12} for <u>C. botulinum</u>. At temperatures other than 121.1^oC, the lethality, F_T, is calculated by

$$F_T = F_o \times 10^{(T-121.1)/z} , \qquad (eq. 1)$$

where z is the increase in temperature that decreases the thermal death time by a factor of 10 (z=10^oC for <u>C. botulinum</u>). In a thermal processing experiment under changing temperatures, the recording instrument (Digistrip) is programmed to calculate incremental F_o values

Broccoli, control

Broccoli, heated

Time, min
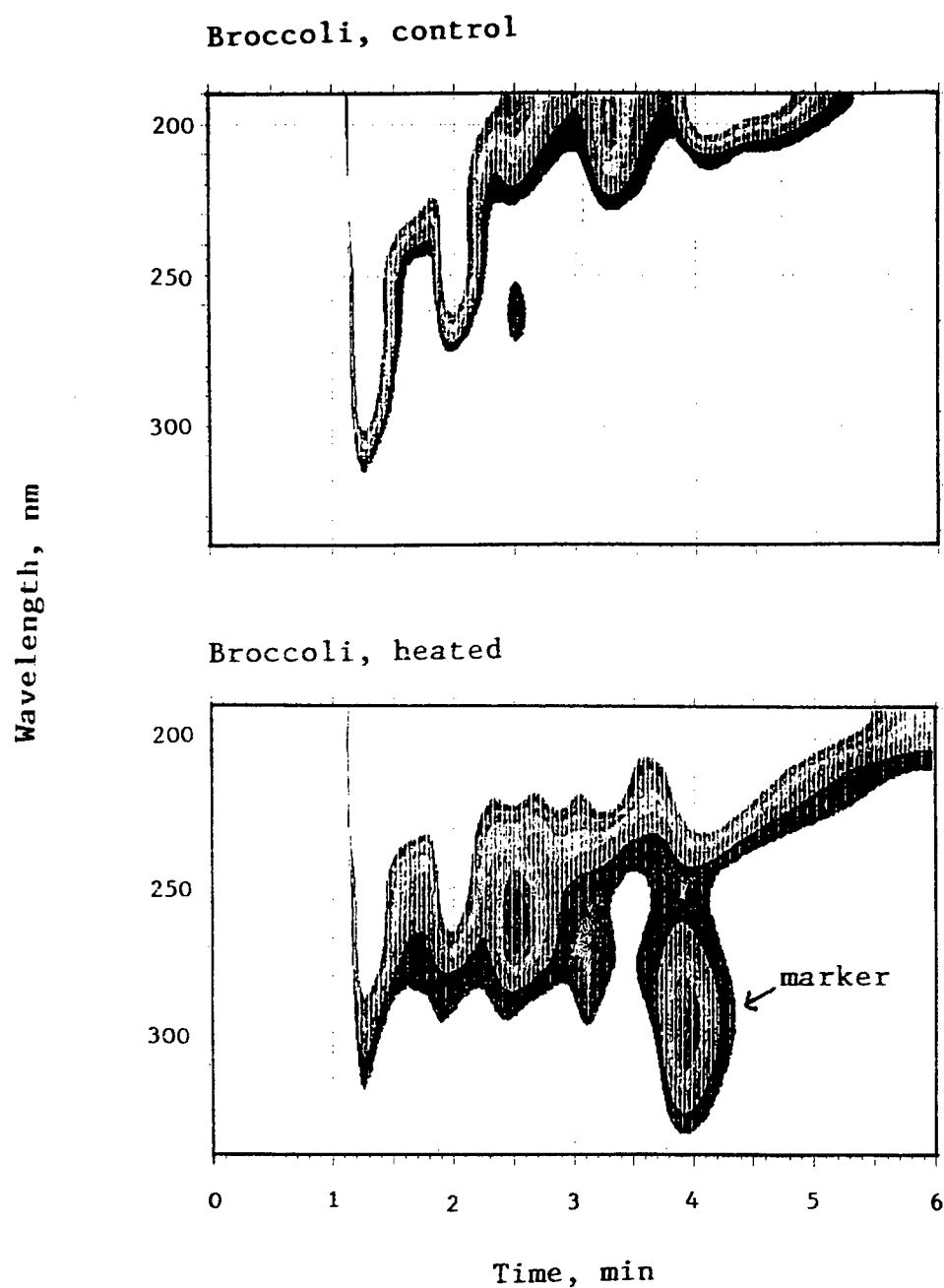
Fig. 1 Contour diagram corresponding to the
spectrochromatogram for control broccoli
and heated broccoli

according to eq. 1 from the measured temperature and print out accumulated F_o values. The accumulated F_o value is the index of sterility at the site of temperature measurement. F_o for C. botulinum is 2.5 min. A good correlation with F_o is the key to adopting any chemical marker, since the objective is to assess F_o from measurement of the marker concentration.

In order to determine if a correlation exists between the marker concentration and F_o values, experiments were conducted with a puree of potato in brine prepared by blending 160 g cut potatoes with 120 ml hot brine (3.4% sugar, 1.7% salt). Ten 20 ml glass vials were filled with the puree, sealed with rubber septum and aluminum cap, and placed in a laboratory oven preheated to 130°C. Two of these vials were used for temperature measurements. One thermocouple was pushed through the rubber septum and placed at the center of one vial. Another thermocouple was placed in another vial midway between the wall and its center. The temperature and the accumulated F_o were printed out every minute by the recorder. The average of the F_o measurements at two locations was used for interpretation of data. At certain predetermined F_o values, a vial was withdrawn from the oven and quickly cooled with cold water. The content of each vial was homogenized with 5-fold excess water and the extract was filtered through a 0.45 μm membrane filter. The marker (M-1) concentration in the filtrate was determined by the AEC-PDA system.

As shown in Figure 2, the average temperature at two locations within the vial increased slowly toward 130°C. As the temperature increased, the accumulated F_o increased rapidly. The marker was undetected in the unheated sample, but its concentration in the heated samples increased as the temperature and accumulated F_o increased. More importantly, the observed marker concentration showed a good overall correlation with F_o values. Although the data need to be supplemented at low F_o values, the results indicate that M-1 has a great potential as a time-temperature integrator of thermal processing.

The fact that the correlation is valid at high F_o is important. Since the attained lethality must exceed a minimum required value (usually F_o of 6 for low acid foods), it is desirable to optimize the processing parameters under slightly overprocessing conditions. In a direct microbiological procedure, it is difficult to recover any survivors from an inoculum of 10^6 spores after F_o of 6 min has been exceeded at the center of a particulate, so the extent of overprocessing cannot be estimated from the postprocess assays. The same is often true even with the more resistant thermophilic spores.[3] In contrast, the extent of overprocessing could be easily estimated using the marker, because the marker concentration increases steadily and maintains a good correlation with accumulated F_o. This observed correlation indicates that the marker is a time-temperature integrator over a wide range of F_o values.
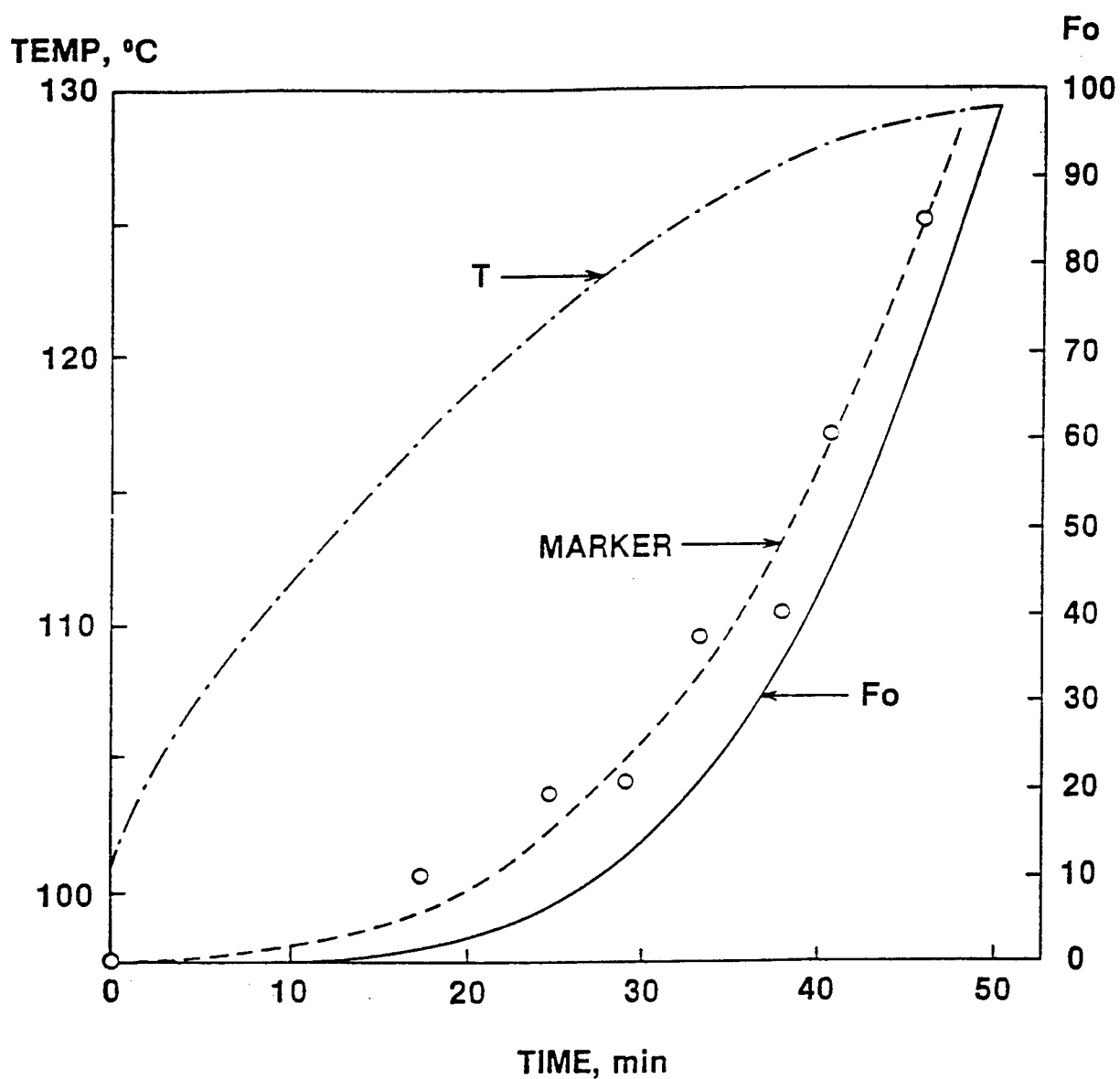
Fig. 2   Correlation of the observed marker
concentration with the lethality value,
F_o, calculated from the temperature (T).
T is the temperature-time profile. The
marker concentration is on an arbitrary
scale (not shown).

PRELIMINARY MARKER CHARACTERIZATION

The mode of the reaction leading to the marker compound has a significant bearing on the usefulness of the compound as a chemical marker for thermal processing. The thermal death of bacterial spores.can be represented as a simple first-order process as follows:

$$\text{LIVE SPORE} \xrightarrow{\text{heat}} \text{DEAD SPORE} \qquad \text{(eq. 2)}$$
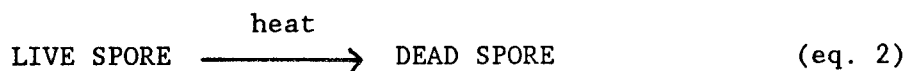
A first-order reaction for the production of the marker resembling the thermal death of bacteria could be written as follows:

$$\text{PRECURSOR} \xrightarrow{\text{heat}} \text{MARKER} \qquad \text{(eq. 3)}$$

In order to investigate the possible correspondence of the two processes, the mode of formation of the marker from its precursor was examined.

The aqueous extract from heated broccoli was fractionated on a gel filtration column (1.0 x 42 cm) packed with Bio-Gel P-2 (molecular weight cut-off: 2,000 daltons) at a flow rate of 10 ml/h. A 1.7 ml volume was collected per fraction. The collected fractions were analyzed for the marker, M-1, by the AEC-PDA system and the marker was found in fraction 24. The gel filtration column was calibrated with riboflavin phosphate (MW=455, fraction 12) and tryptophan (MW=202, fraction 32). The molecular weight of M-1 in fraction 24 was estimated as approximately 280.

In order to investigate the precursor, the aqueous extract from unheated broccoli was fractionated on the same gel filtration column and the collected fractions were heated in sealed glass tubes. When the heated fractions were analyzed for M-1, only fraction 22 (MW $\approx$ 300) yielded the marker. This observation suggests that the net effect on the precursor of the thermally-induced reaction is the removal of a small functional group.

When the aqueous extract was heated in sealed capillary tubes, an exponential increase in the marker concentration up to a limiting value was observed. The limiting concentration presumably reflects depletion of the precursor. When the logarithm of the difference between the limiting value and the observed value was plotted against heating time, a linear relationship was obtained. This first-order behavior is consistent with the proposed reaction in eq. 3 whose rate is solely dependent on the precursor concentration. This mode of reaction resembles thermal death of microorganisms represented by eq. 2 and supports the proposition that the compound is a potential indicator of sterility. Purification of M-1 is in progress, and elucidation of the structure by infrared spectroscopy and mass spectrometry will follow.

TEST APPLICATIONS

Potato Cooked at Low Temperature

The occurrence of the marker, M-1, was investigated in potato cubes (approximately 1 cm cube) in home-cooked beef stew. The stew was simmered for about 30 min, and the potato cubes were frozen for analysis of the marker. Each partially thawed potato cube was sliced with a razor blade. Each surface slice or remaining center portion was placed in a test tube containing five-fold excess water, and the content vigorously mixed using a vortex mixer. The aqueous extract was filtered and injected into the chromatograph for determination of M-1 by the AEC-PDA system. The marker was present at a significant concentration in the surface slice of the potato cube, but it was barely detectable from the center portion. These qualitative results indicate that the temperature gradient within the potato cube established under the cooking conditions is reflected in the concentration differential of the marker, which makes M-1 a useful indicator of thermal processing.

Aseptically Processed Chicken Meat

Chicken meats in chicken stew aseptically processed at an industrial facility (Alfa-Laval, Newburyport, MA; courtesy of Dr. Margaret Driver) were tested for occurrence of the marker, M-2, at different locations within the meat piece. Chicken meats were frozen with ice immediately following aseptic processing at 129°C at two flow rates, 500 lb/h and 1,300 lb/h. The meats, which were of irregular shape (approximate dimension, 1.5x2.5x3 cm), were divided into surface, middle, and center portions of approximately equal thickness with a razor blade, and these portions from several meat pieces were pooled together. Each aqueous extract from these meat portions was concentrated by freeze-drying and then redissolved in a small volume of 5mM sulfuric acid solution. The concentration of M-2 was determined by the AEC-PDA system as described above.

Several important features are evident in the results summarized in Figure 3. First, a definite concentration gradient exists for the marker, decreasing gradually in the direction of heat transfer from the surface to the center. This concentration gradient presumably reflects the temperature gradient within the meats. The concentration gradient was observed at both flow rates. Second, the marker concentration is higher at the lower flow rate, which corresponds to a longer residence time. As a matter of fact, the concentration of M-2 at the surface was inversely proportional to the flow rate (or directly proportional to the residence time), which strongly suggests that the marker is a good indicator of heat exposure. Third, the marker concentration at the center of the meat processed at the lower flow rate (500 lb/h) is higher than that at the surface of the meat processed at the higher flow rate (1,300 lb/h). This observation suggests that lethality at the center of a food particulate will be determined more by external parameters such as
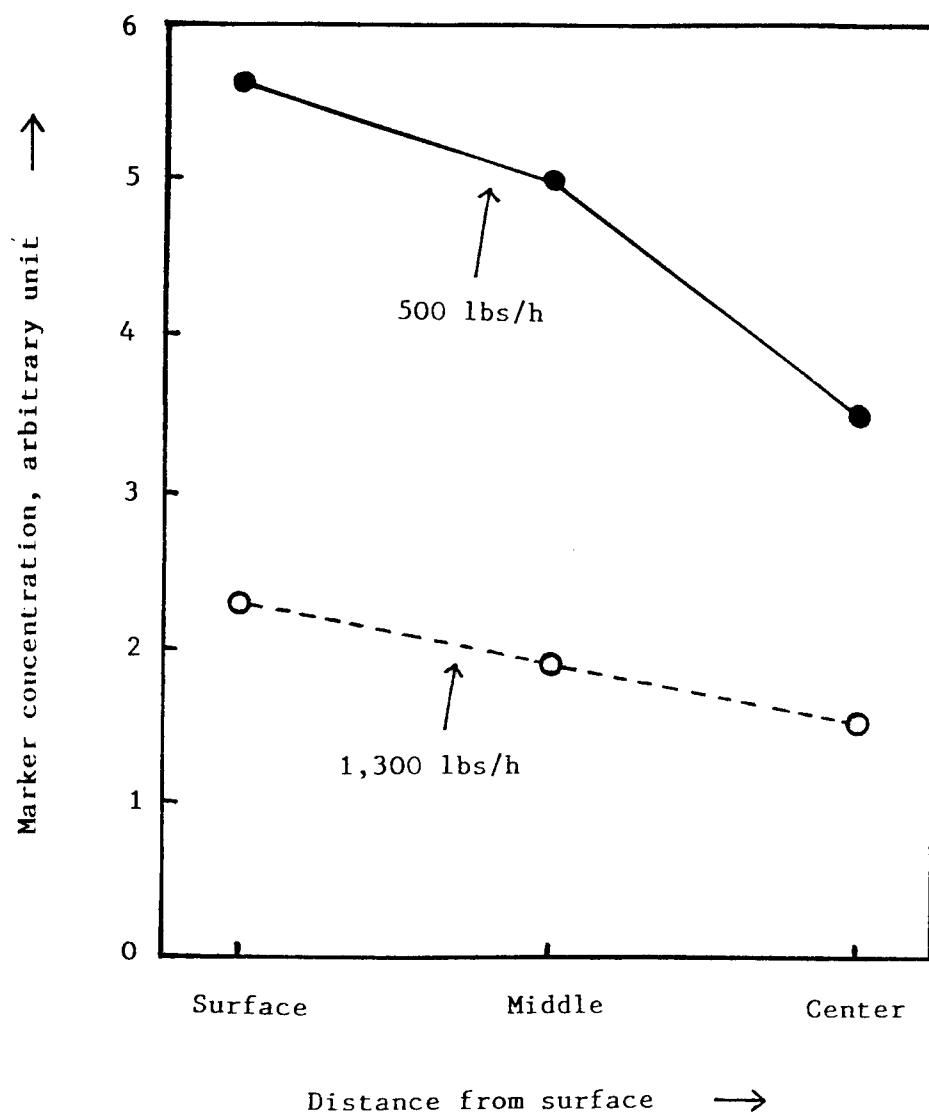
Fig. 3 Concetration of marker, M-2, at surface, middle, and center portion of chicken meat aseptically processed at 129°C

temperature and flow rate than by heat transfer characteristics, which are inherent in the food system and depend on composition and microstructure.

It should be noted that the marker was observed at a significant concentration at the center even at the higher flow rate. A correlation such as shown in Figure 2 is needed for M-2 at low F_o values in order to translate this concentration to F_o. The precursor concentration or limiting marker concentration in the meat also needs to be determined for accurate assessment of F_o. Nevertheless, the above observations indicate that the compound, M-2, is a good candidate marker for meats.

CONCLUSION

Two compounds were detected and separated for the first time that appear to be extremely useful chemical indicators of time-temperature exposure at different locations within a food particulate. They are easy to detect with a high sensitivity. Once the precursors are identified and the kinetic parameters are measured, it will be possible to accurately determine the F_o at different locations within a food particulate from the observed marker concentration and limiting marker concentration in heated food.

REFERENCES

1. Dignan, D.M., Berry, M.R., Pflug, I.J., and Gardine, T.D., Food Technol., 43(3), 118 (1989).

2. Sastry, S.K., Li, S.F., Patel, P., Konanayakam, M., Bafna, P., Doores, S., and Beelman, R.B., J. Food Sci., 53, 1528 (1988).

3. Hinton, Jr., A., Driver, M.G., Silverman, G.J., and Taub, I.A., Activities Report of the R&D Associates, 41(1), 39 (1989).

4. Heldman, D.R. Food Technol., 43(3), 122 (1989).

5. Chandarana, D.I., Gavin III, A., and Wheaton, F.W., Food Technol., 43(3), 137 (1989).

6. Mulley, E.A., Stumbo, C.R., and Hunting, W.M., J. Food Sci., 40, 993 (1975).

Sub-nanosecond Risetime High Power Photoconductive GaAs Switch
and Its Transient Electric Field Profiles

*A. Kim
R. Zeto
R. Youmans
M. Weiner
U.S. Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

I. INTRODUCTION

The interest in high power pulse generation using photoconductive switches stems from potential applications such as impulse and microwave generation [1, 2, 3]. In particular, the generation of high amplitude pulses with sub-nanosecond pulsewidths has attracted interest because of their wide bandwidth properties and their potential as a wide bandwidth RF source [4].

The two critical elements of high power pulse generation with sub-nanosecond pulsewidth are the switch and the energy storage element. To generate such pulses, the switch must transition from the high resistivity state to the conducting state in a sub-nanosecond time frame. Photoconductive switches based on semi-insulating GaAs material have been observed to have the necessary switching speed. Studies have been conducted investigating the fast switching mechanism and the external circuit effects, which can attenuate output pulse amplitude and slow down pulse risetime. Attention was given to techniques in which the energy storage element delivers the narrow pulses, via the photoconductive switch, to the load impedance.

A great deal of effort was devoted to the photoconductive switch design. The objective was to achieve high voltage capability, good effective quantum efficiency, and fast switching. A photoconductive bulk GaAs switch with two opposite gridded electrodes [5] was designed, fabricated, and tested at the ET&D Laboratory. In this switch geometry, the laser light is introduced parallel to the applied electric field. This unique switch geometry enables one not only to extend the effective gap distance but also to generate a symmetrical distribution of the photon generated carriers throughout the entire active switching area. Experimental efforts were conducted to investigate these switches. In particular, the dynamic electric field profiles were obtained using an optical probe technique.

In general, two techniques are used to deliver the electrical pulses to the load impedance via a photoconductive switch. The first technique uses the recombination property of the semiconductor material. If the recombination time is small (compared to the pulsewidth of the stored electrical energy), then narrow pulses will be produced when correspondingly narrow light pulses are utilized. The pulses generated with this technique using GaAs, however, may have a long recovery at high voltages, because of the lock-on phenomenon [6, 7].

In the second approach, the output pulsewidth is controlled by the energy storage element. Previous work succeeded in achieving reduced pulsewidths, but the voltage amplitude at the load was limited to about half of the charging voltage. In the present work the unique concept of a radial transmission line was employed for the energy storage element. This approach produced a large capacitance for the given finite physical dimensions, and a voltage amplitude enhancement was achieved because of the impedance transformation (from low impedance to high impedance) as the wave in the radial transmission line proceeds towards the center of the structure. Another important advantage is that a wafer scale integrated pulser can be fabricated on the large diameter wafer (to be discussed). In addition, the radial structure allows one to incorporate an antenna pattern on one of the faces of the radial structure so that the functions of switch, energy storage, and the antenna can be placed on a single GaAs wafer.

II. SWITCH DESIGN

A. Switch Design

The voltage capability of the photoconductive bulk GaAs switch increases with the gap distance between the two electrodes. However, making switches with a long gap distance becomes complicated due to the limit of laser light penetration depth. The amount of laser energy needed to trigger the switch is another factor limiting switch gap distance. In an effort to achieve high voltage capability, good quantum efficiency, and fast switching time, a photoconductive GaAs switch with two opposite gridded electrodes was designed. Since light is introduced from both sides, a switch design having a much wider gap distance was realized. As well as a more uniform carrier distribution throughout the gap. The conceptual diagram of the photoconductive switch triggered from two opposite sides is given in Figure 1.

A previous investigation [8] demonstrated that a gridded silicon PIN diode had good efficiency and voltage hold-off. Experimental work was conducted in similar GaAs switches to find the optimum relationship between laser energy coupling efficiency and electrical field uniformity.
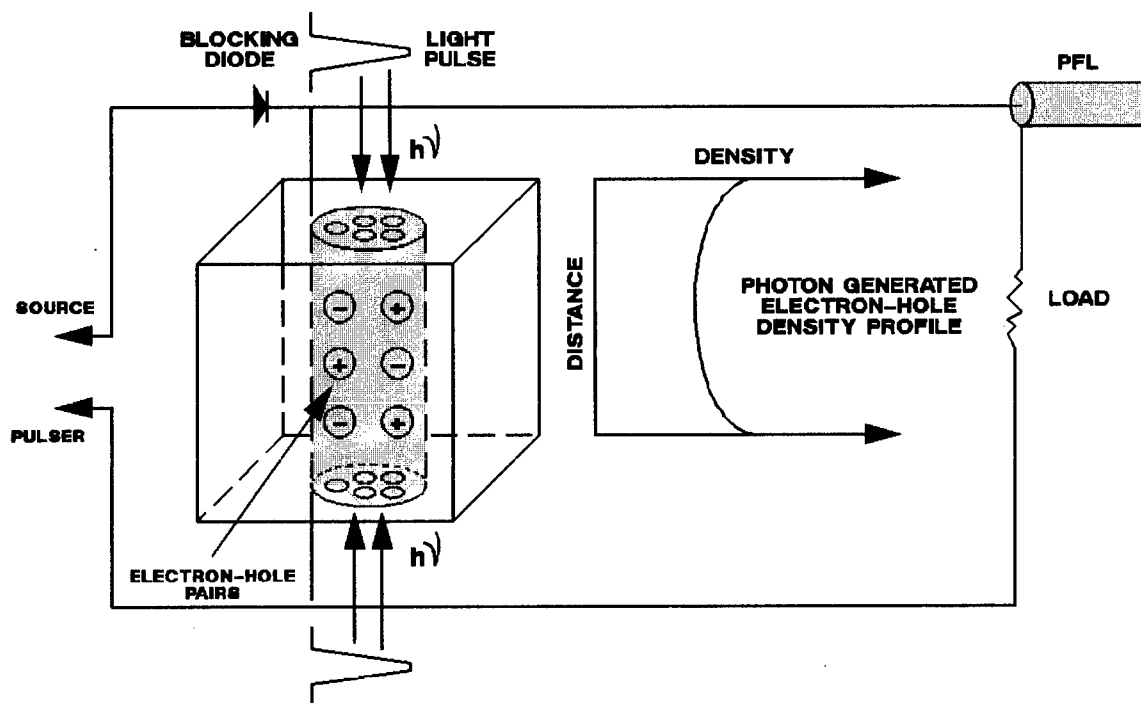
FIGURE 1. THE CONCEPTUAL DIAGRAM OF HIGH POWER PULSE GENERATION USING PHOTOCONDUCTIVE SWITCH

Four different aperture electrodes, which gave 20%, 40%, 60%, and 100% openness ratios, were designed. The aperture separation in the electrodes was fixed at 15 microns, except for the 100% openness ratio. It was found that the 40% and 50% openness ratios had the best efficiencies [7]. Based on this result, larger, dimensionally scaled bulk photoconductive GaAs switches with the two opposite gridded electrodes, in which the outer diameter of the electrode was 1.0 cm, were designed using a 2.2 cm X 2.2 cm X 1.0 cm, wafer of semi-insulating GaAs material. The matrix of electrode apertures, each 586 micron in diameter and separated by 150 microns, were located within an area of 0.6 cm diameter.

B. Switch Fabrication

Switches were fabricated by Class 100 clean room processing of semi-insulating (100 GaAs wafers from M/ACOM. Substrate resistivity was 4E7 ohm-cm with dimensions of 5.1 cm diameter and thickness between 0.5 cm and 1.0 cm. The wafers were chem-mechanically polished to remove surface damage and provide a planar optical wave front. Two switches, each 2.2 cm on edge, were photolithographically defined on each wafer using orthogonal saw cuts as a reference for alignment of the photomask on each face.

The wafers were ultrasonically cleaned in acetone and microclean solution, and successively rinsed in deionized water, acetone, methanol, and deionized water. After blow-drying in nitrogen, the wafers were dehydration baked in an oven at 150^0C. AZ 1350J positive photoresist was spun on a wafer face using HMDS (Hexamethyldisilazane) as an adhesion promoter. After a 45 min. oven bake at 90^0C, the process was repeated on the other face. Each face was successively exposed in a Kasper System 3001 Aligner, after which the wafer was swirled in chlorobenezene, developed, rinsed in deionized water, and plasma cleaned. Each face of the wafer was metalized with 50 Å nickel, 300 Å germanium, 600 Å gold, 1000 Å silver, and 1000 Å gold. Photoresist and excess metal were lifted-off by ultrasonic immersion in acetone, leaving gridded aperture electrode patterns. The wafers were rinsed in methanol, dried, and the photoconductive switches were sawed from the wafer. The electrodes were augmented with indium metal to make good contact with the external circuit.

III. INVESTIGATION OF THE SWITCHING MECHANISMS

An investigation was conducted to understand the effects of the biasing voltage and the laser light illumination on the risetime of the produced electrical pulses was investigated using a Q-switched Nd:YAG laser, which produces a 12 ns wide pulse. The amount of input laser energy was kept constant throughout the entire sequence of measurements.

The biasing voltage amplitudes were increased from 5 kV to 12 kV with a 1 kV incremental voltage. The experimental results at 14 kV are shown in Figure 2.

Collapse of the electric field in the GaAs switch under laser light illumination was investigated collaboratively with an optical probe technique [9] developed at the University of Rochester/Laboratory for Laser Energetics. After the 1.0 cm long photoconductive GaAs switches were cut in half, parallel to the direction of the electric field, a LiTaO3 crystal was put over the entire active area of this switch. The electric field of the GaAs has a component that extends above its surface, and the tangential component of the electric field is continuous at interface. The electric field inside the LiTaO3 interacts via the electro-optic coefficient, so that its produces a change in the briefringence of the crystal. Hence, the LiTaO3 couples the surface field with the polarization of an optical pulse. When imaged onto a two-dimensional detector array, this technique produced snapshots of the surface electric field profiles with 200 ps time resolution and 3 um spatial resolution. The electric field profiles dependence on the illumination is given in Figure 3.

IV. DESCRIPTION OF SUB-NANOSECOND DESIGNS

A. Hybrid Pulser Using a Radial and a Quasi-radial Transmission Line

The hybrid pulser [10] is one in which the dielectric medium used to store energy is different from the material used as the switches (i.e., the semiconductor). The concept of the hybrid pulser using a radial transmission line as the energy storage element was experimentally verified using a hybrid design shown in Figure 4. The outer and inner diameters of a radial transmission line, which is made of an acrylic materials with a dielectric constant of 2.3, were 25 cm and 5 cm, respectively. A photoconductive GaAs switch, 0.5 cm thick and 2.2 cm wide, was situated at the center of the energy storage medium. The GaAs switch at the center makes contact with both the high voltage electrode of the radial line and the inner conductor of the output coaxial line. Laser light is introduced in the center of the switch thru grid apertures in the switch electrode. Photon generated carriers induce a change in the switch from a high resistivity state to a high conductivity state. The electric field sweeps the electron-hole pairs across the switch active area, so the electrical energy stored in medium between conducting plates is delivered to the coaxial load. The inner diameter of the radial transmission line is designed to match the impedance of the output coaxial line (typically 50 ohms).
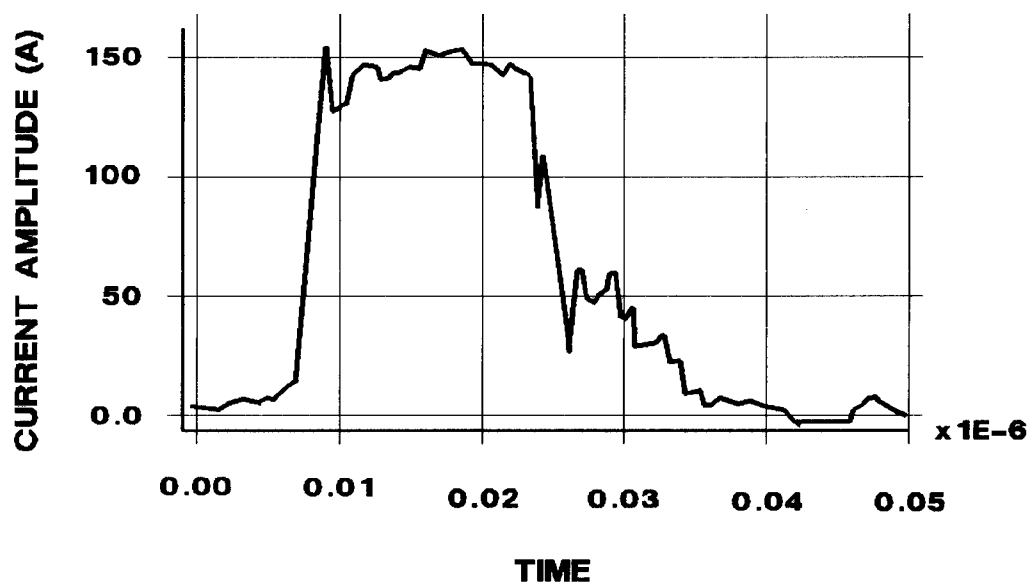
FIGURE 2. CURRENT WAVEFORM OBTAINED FROM 14KV BIAS VOLTAGE.
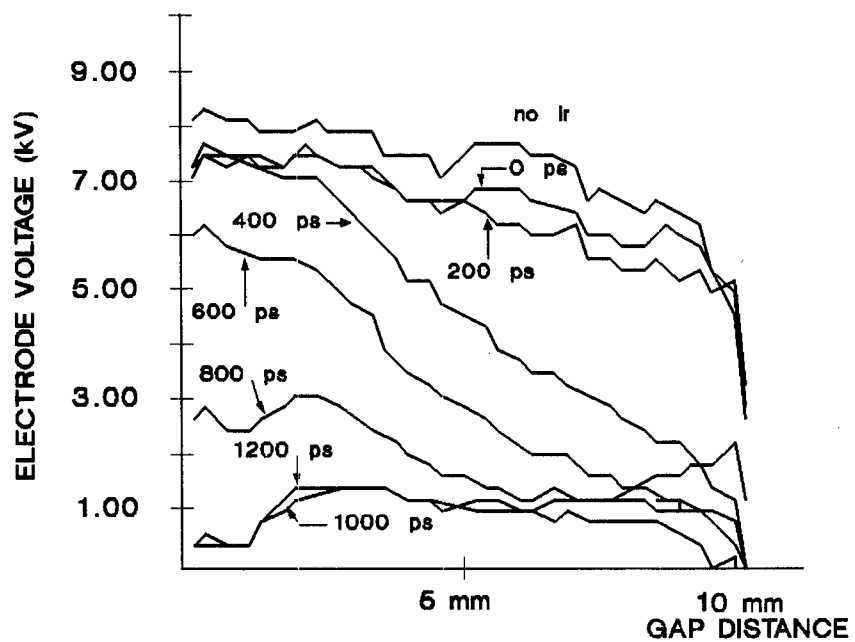
FIGURE 3. ELECTRIC FIELD PROFILE OBTAINED WITH BOTH SIDE ILLUMINATIONS.
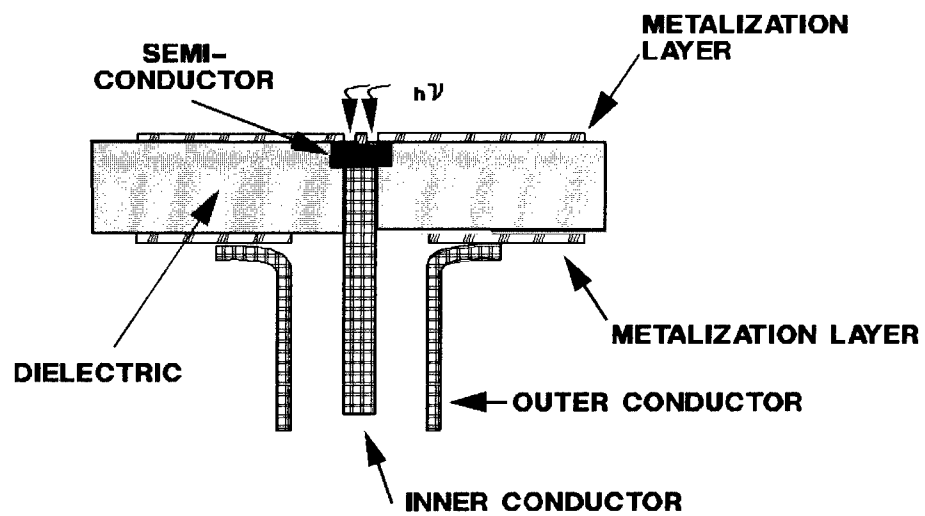
FIGURE 4. HYBRID PULSER UTILIZING A RADIAL TRANSMISSION LINE.

If one neglects a host of factors, such as the reflection in the radial transition regions and dispersion effects, simple approximations for the voltage transformation can be made. The output voltage amplitude can be estimated as follows:

$$V_{out} = [V_o Z_o/(Z_{IR} + Z_o + R_s)] \times [Z_{IR}Z_{OR}]^{1/2} \qquad (1)$$

where Z_{IR}, Z_{OR} are the impedance in the radial line at the inner and outer diameter respectively, Z_o is the output impedance of the coaxial line, R_s is the semiconductor resistance, and V_o is the bias voltage. If one assumes instantaneous switching and a matched condition at the beginning, the output voltage will immediately jump to half of the biasing voltage, and then ramp up to the peak voltage given in equation (1), in a time equal to the one way transit time in the radial transmission line. A similar decline in voltage then follows.

The design of the hybrid pulser using a quasi-radial transmission line was accomplished by replacing the upper electrode of the previous design with six parallel microstrip lines. The width of each microstrip conductor line was 2.5 cm. The height between electrodes was 3 cm. Although each stripline represents a uniform transmission line having a constant impedance, the coupling between microstrip lines causes an impedance transformation. As the wave starts to move toward the center, the wave sees an increasing impedance. However, the impedance transformation of a quasi-radial transmission line is significantly smaller, compared to that of the radial transmission line. On the other hand, it is expected that the quasi-radial transmission line will have better risetime and pulse width characteristics, but with lower amplitude gain compared to the radial transmission line.

B. Integrated Pulser

The integrated pulser is the one in which the energy storage element and the switch are fabricated from the same material. In this device, a pair of circular electrodes are fabricated on two opposite sides of a high dielectric constant semiconductor wafer with the switch region located in the center. The electrical energy is stored in the semiconductor material itself, and the switch is activated with optical energy. It is necessary to use a high resistivity semiconductor material so the effect of a dark current can be minimized at high biasing voltage. A semi-insulating GaAs material is chosen because its dark current is negligible for microsecond wide bias pulses. The reason for the p-i-n structure of this pulser is to enhance voltage capability.

Another unique feature of this design is the possibility of the switch to self extinguish. As the photon generated carriers sweep across the active area of the switch, conducting plasma will spread out and eventually reach the outer conductor of the output coaxial cable. since this process results in a complete short circuit of the output, electrical pulses having very fast falltime are produced regardless of any energy associated with impedance mismatch in the circuit.

V. EXPERIMENTAL SET-UP FOR THE HYBRID PULSER

A schematic of the test set-up is shown in Figure 5. A mode-locked Nd:YAG laser (Quantel YG-501C) generates 1.06 microns wavelength optical pulses having a pulsewidth of approximately 75 picoseconds. After attenuation of the laser output, the laser light was coupled into a fiber optic bundle and delivered to the switch. About 1.0 mJ of optical energy is illuminated to each gridded electrode.

Initially, the hybrid pulser is pulse charged using a combination of SCR and step-up transformer circuit. After completion of the charging process, the charging circuit is isolated from the remaining circuit. As soon as the laser is triggered, photon generated carriers were produced and the switching process was initiated. A 6 GHz Tektronix 7250 transient digitizing oscilloscope was used to measure the sub-nanosecond risetime pulses. The generated pulse from the hybrid pulser was coupled to a 50 ohms coaxial cable, and the pulse amplitude was attenuated by 2000 to 1 ratio, using high voltage, wide bandwidth Barth and General Radio attenuators. The electrical pulses produced by the hybrid pulser, using a radial and a quasi-radial transmission line at 9 kV, are shown in Figure 6. The actual bandwidth of the measurement was estimated at about 4 GHz due to the delay line which had only a 4 GHz bandwidth.

VI. EXPERIMENTAL RESULTS

A. Switch Properties

Initial switching results were obtained with a Q-switched Nd:YAG laser. The risetime of the obtained pulses became faster as the biasing voltage increased. Although the triggering optical pulsewidth was approximately 12 ns, the risetime of the electrical pulses at 10 kV became almost 1 ns, limited by the external circuit. Additional investigations were conducted on the effect of the light intensity. The most significant effect of the light intensity was related to the pulse amplitude. It was found that the light intensity had a negligible effect on the switch risetime as long as the light intensity was above a threshold energy of approximately 0.1 mJ.
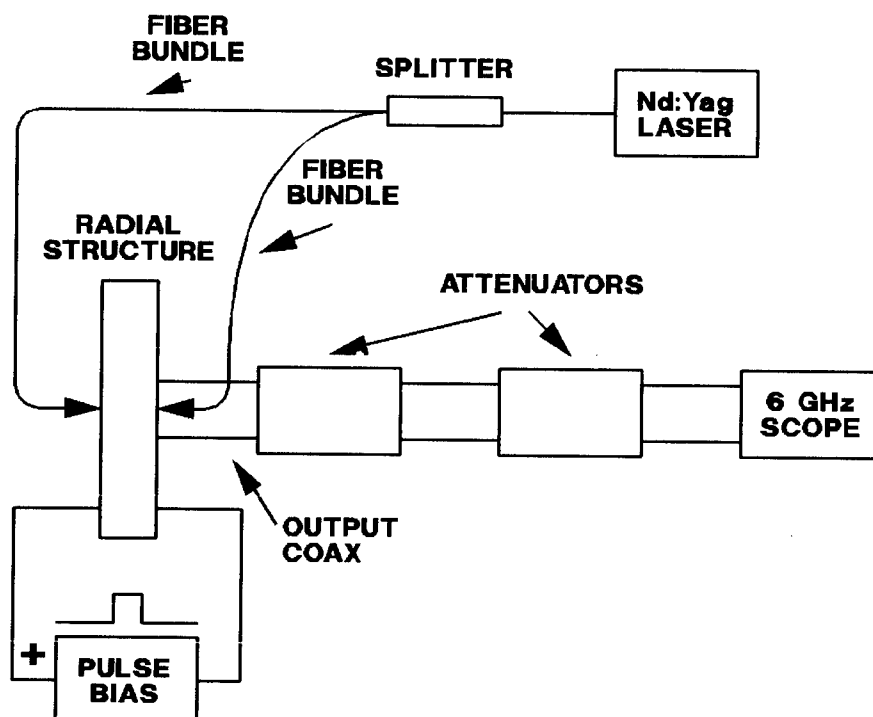
FIGURE 5. A SCHEMATIC OF THE TEST SET-UP FOR
HYBRID PULSER.

The effect of the laser illumination on the risetime of the output pulse also was investigated using an optical probe technique. The results obtained with a constant input laser light showed that the two-sided illuminations gave faster risetime compared to the one-sided illumination. In addition, as shown in Figure 3, the results obtained with the optical probe technique showed that the electric field collapsed much faster when the lights were illuminated from both sides. The experimental results generally agree with the prediction based on the switch geometry. It is expected that advances in the design of this structure will produce extremely high power photoconductive switches, operating possibly in a non-destructive avalanche mode of operation.

B.   Radial Line Results

As shown in Figure 6, the radial transmission line produced higher voltage amplitudes compared to the quasi-radial transmission line. The observed voltage gains are about 2.2 and 1.5 for the radial and quasi-radial lines respectively. The output voltage gains are in accord with values based on the estimated impedance levels of the transmission line. As the voltage increases, the falltime of the pulses was observed to decrease. This behavior may be connected with non-linear behavior in the switch. A complete circuit analysis, which takes into account the radial impedance transformation, the RF discontinuities, and the transient properties of the switch for the given experimental configuration, must be examined to identify the anomalous falltime.

VII. CONCLUSIONS

Photoconductive GaAs switches with gridded electrodes were designed and fabricated, and have demonstrated voltage capability over 45 kV. A sub-nanosecond transient switching mechanism was investigated using an optical probe technique. Parametric studies were conducted to investigate the effects of the biasing voltage and the transmission line impedance. Sub-nanosecond pulsers, which incorporate a photoconductive GaAs switch into a radial and a quasi-radial transmission lines, have been investigated. Peak voltage amplitude close to 10 kV was produced with 8 kV biasing voltage. The concept of the integrated pulser, in which the energy storage element and the light activated switch are fabricated in one wafer, has been examined. This totally integrated pulser is expected to have advantages in speed, faster risetime, and compactness.

In particular, the optical probe technique is expected to have an enormous impact on high power photonconductive switching development. With advances in this technique, one will be able to measure important parameters needed to characterize semiconductor switching properties, such as avalanche breakdown, ohmic contacts, localized field enhancement, etc.
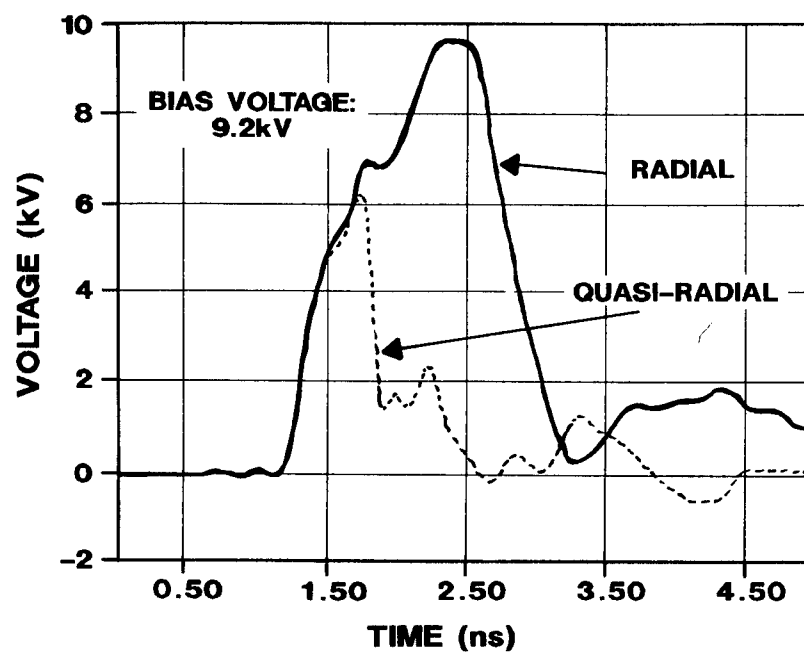
FIGURE 6. VOLTAGE WAVEFORM OBTAINED FROM 9.2kV BIAS VOLTAGE

REFERENCES

[1] G. M. Loubriel, M. W. O'Mally, and F. J. Zutavern, "Toward Pulsed Power Uses for Photoconductive Semi-conductor Switches: Closing Switches," Proceedings of the 6th IEEE Pulsed Power Conference, pp 145-148, June 1987.

[2] W. C. Nunnally and R. B. Hammond, "80-MW Photoconductor Power Switch," Appl. Phys. Lett., Vol. 44, pp 980-982, May 1984

[3] A. Kim, M. Weiner, R. Youmans, P. Herczfeld, and A. Rosen, "High Power RF Generation with Optically Activated Bulk GaAs Devices," 1989 IEEE MTTS-S International Microwave Symposium Digest, pp 1071-1074, May 1988.

[4] W. B. Scott, "UWB Radar has Potential to Detect Stealth Aircraft," Aviation Week & Space Technology, pp 38-41, Dec 1989.

[5] A. Kim, M. Wade, R. Youmans, M. Weiner, and R. Zeto, "Bulk GaAs Photonic Devices with Two Opposite Gridded Electrodes," Proceedings of the 7th IEEE Pulsed Power Conference, pp 430-432, June 1989.

[6] M. S. Mazzola, K. H. Schoenbach, V. K. Lakadawala, R. Germer, G. M. Loubriel, and F. J. Zutavern, "GaAs Photoconductive Closing Switches with High Dark Resistance and Microsecond Conductivity Decay," Appl. Phys. Lett., Vol. 54, pp 742-744, Feb. 1989.

[7] A. Kim, M. Weiner, M. Dornath-Mohr, M. Wade, R. Youmans, R. Zeto, G. C. Vezzoli, "Photonic Switching and Lock-on Phenomenon in Single Crystal Semi-insulating GaAs," Proceedings of the SPIE Conference, Vol. 993, pp 137-140, Sep. 1988.

[8] "Optically Activated Switch for MW Wave and Laser Transmitters," Final Report, U.S. Army Contract DAAL01-85-C-0421, SRI, Princeton, New Jersey, 1988.

[9] W. R. Donaldson, L. Kingsley, M. Weiner, A. Kim, and R. Zeto, "Electro-optic Imaging of the Internal Fields in a GaAs Photoconductive Switch," Submitted to J. Appl. Phys., Dec. 1989.

[10] A. Kim, M. Weiner, R. Zeto, R. Youmans, L. Jasper, and B. Lalevic, "Photoconductive Sub-nanosecond Pulse Generation Utilizing Radial Transmission Lines," Submitted to Tran. IEEE Ed., Mar. 1989.

The Effect of Hydrogen, Boron, Carbon, Phosphorus and Sulphur on Intergranular Cohesion in Iron

Genrich L. Krasko, Dr.
Metals Research Branch, U. S. Army Materials Technology Laboratory,
Watertown, MA 02172-0001

Introduction

"First principles" design of materials is a pipedream of many materials scientists. The electronic theory of the solid state, since its introduction in the beginning of this century, has always had this challenging goal as one of its ultimate objectives. Formidable as it is, this problem is of utmost importance for both the future intelligent development of new materials and the improvement of the quality of existing ones.

The recent progress in high speed computers and software has stimulated a more aggressive approach by physicists and materials scientists in their attempts at first principles understanding of not only the traditional electronic properties of materials, but their mechanical properties as well. A new family of computationally efficient methods for electronic band structure and total energy calculations enables such delicate properties as cohesive energy, surface and grain boundary energies, magnetic properties, etc., to be calculated with the necessary precision.

In spite of the complexity of parameters effecting mechanical behavior and difficulties in characterization of the corresponding structures on microscopic levels, important trends in mechanical properties can now be correlated with definite features of electronic structure. The latter is extremely important and has been applied to this theoretical alloying study of the effects of impurity solutes on steels.

In metallic alloys (including technologically important materials such as steels), there are many ways, and, in fact, a significant reserve, of known mechanisms for increasing strength - the ability to withstand plastic deformation. A major deterrent for these mechanisms is that the strengthened material usually experiences a reduction in ductility and fracture resistance. This is particularly acute in ballistically resistant steels where increased strength and hardness is accompanied by reduced material plasticity.

The reduced cohesion of grain boundaries (GBs) is often the controlling factor limiting the ductility of high strength metallic alloys[1]. This is particularly so when an environmental interaction is involved, as in the case of intergranular hydrogen stress corrosion cracking, or in the case of temper embrittlement of ultrahigh strength steels.

Intergranular embrittlement is typically associated with prior segregation of impurities toward the GBs[2-5]. Impurities present in bulk concentrations of 10^{-3}-10^{-4} atomic percent can result in a dramatic decrease in plasticity. Both a simple estimate and numerous
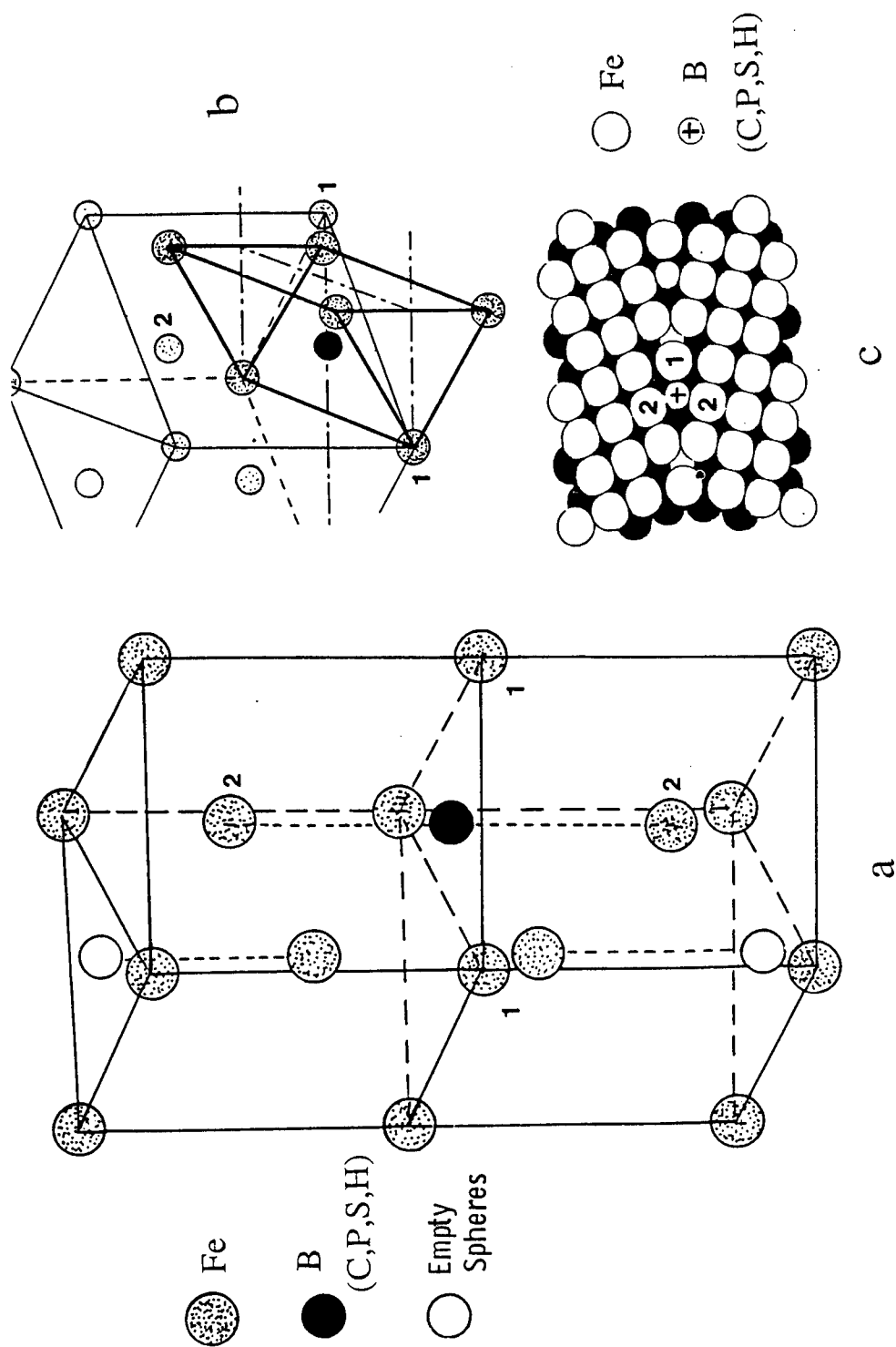
Fig.1 a) The 8-atom hexagonal supercell representing the (111) Σ3 GB in Fe. The Fe-1 atoms lie in the GB plane, while the Fe-2 atoms are situated across the GB; b) The trigonal prism configuration of Fe atoms in the GB core; c) The Fe-1 and Fe-2 atoms on a true GB

experiments show that these minute amounts of impurities can completely "saturate" the GBs for typical grain sizes.

Recent modelling of intergranular crack propagation [6] predicts a correlation between the embrittling potency of impurities and the difference between their segregation energies at free surfaces and GBs. Solutes with a lower energy at free surfaces reduce boundary cohesion, while those with a lower energy at GBs are cohesion enhancers. In the case of Fe and its alloys, sufficient thermodynamic data are available to demonstrate the plausibility of such a description of intergranular embrittlement [6].

A microscopic basis for multicomponent alloy design for enhanced intergranular cohesion can be sought from an examination of the electronic origins of the relative energies of impurities in the GB and the corresponding free surface environment. The present study of impurities in Fe GBs is a first step toward that goal, which would mean for Army applications the availability of higher strength armor and structural steels with enhanced ductility.

Attempts at qualitatively explaining GB decohesion on the electron-atom level go back to the 1960s [7]. Quantitatively, the electronic structure calculations treating small atomic clusters as a simple model of GBs [8-13] shed some light on intergranular cohesion in metals. First-principles methods using two-dimensional geometries were applied to study GBs in Fe[14] and Ni [15].

The results accumulated thus far are somewhat controversial. In Refs. [8,9,11,12,14] it was shown that impurities on the GB result in weakened bonds between the host atoms. Results of the present study demonstrate the weakening of host-impurity bonds. A similar effect was found in Ni [15].

In the present paper the result of the first self-consistent spin-polarized calculations of electronic structures of B, C, P and S in an atomic environment typical of a GB in Fe will be reported. Molecular dynamic simulations[16] show that an interstitial impurity atom, such as B, is likely to occupy the center of a capped trigonal prism formed by Fe atoms in the core of the GB. It is interesting that P, a substitutional impurity, also reconstructs the Fe GB configuration into a trigonal prism, occupying its center. The trigonal prism configurations are also found in iron compounds Fe_3B, Fe_3C and Fe_3P [17]. As for H, due to its small atomic size, H probably tends to occupy tetrahedral voids in the GB plane, rather than centers of trigonal prisms. In the present work, however, H has been treated in the same way as the other impurity atoms. The analysis of the structural difference between H and other embrittlers will be left for future work.

For the simplest model of the GB an 8-atom supercell was devised (Fig. 1). The corresponding crystal lattice can be viewed as a succession of hexagonal planes with stacking: BABCACBAB. The two stacking faults, BAB and CAC, imitate two (111) $\Sigma 3$ tilt GBs. One of the GBs is occupied by an impurity atom, while the other is '"clean" (an empty sphere is placed inside the trigonal prism of Fe atoms) .

Originally the calculations were performed for P, and the dimensions of the supercell were chosen so that the volume of the trigonal prism was equal to that of the Fe_3P compound. Subsequent calculations for B, C , S and H used the same supercell. While the supercell for P happened to be quite relaxed (the self-consistent pressure equalled only 38 kbar), for the other impurities the supercell was rather expanded (large negative pressures).The results,

Table 1. Total numbers of electrons, n, their s-, p- and d- components, the charge transfers, Δn, magnetic moments, m, and pressures, P.

atom[a]	s	p	d	n	Δn	m	P(kbar)
Fe BCC[b]	0.648	0.766	6.586	8.000	0.0	2.41	
Fe 1	0.632	0.834	6.680	8.146	+0.15	2.07	
Fe 2	0.718	0.684	6.611	8.013	+0.013	2.68	
B	0.952	1.486	0.086	2.524	-0.48	-0.15	-322.0
B gas	2.0	1.0	0.0	3.0			
Fe 1	0.575	0.776	6.702	8.053	+0.053	1.88	
Fe 2	0.717	0.678	6.621	8.016	+0.016	2.66	
C	1.355	2.309	0.086	3.750	-0.25	-0.12	-344.6
C gas	2.0	2.0	0.0	4.0			
Fe 1	0.673	1.033	7.065	8.771	+0.77	1.12	
Fe 2	0.719	0.693	6.654	8.066	+0.066	2.62	
P	1.185	1.973	0.236	3.394	-1.61	-0.06	37.6
P gas	2.0	3.0	0.0	5.0			
Fe 1	0.622	0.969	7.118	8.708	+0.71	1.088	
Fe 2	0.718	0.685	6.669	8.072	+0.072	2.60	
S	1.422	2.841	0.266	4.529	-1.47	-0.02	-138.4
S gas	2.0	4.0	0.0	6.0			
Fe 1	0.569	0.549	6.537	7.655	-0.34	2.54	
Fe 2	0.720	0.655	6.616	7.990	-0.01	2.62	
H	0.962	0.269	0.055	1.286	+0.29	-0.04	-305.6
H gas	1.0	0.0	0.0	1.0			

[a] WS radii of atoms (a.u.) : Fe-1: 2.733; Fe-2: 2.876; H, B, C, P, S : 2.137

[b] An "equivalent" WS radius of BCC Fe (a.u.): 2.804

however, show an interesting trend that gives a qualitative picture of the effect of the two types of impurities on the intergranular cohesion in Fe.

The Linear Muffin-Tin Orbitals (LMTO) method [18,19] was used. This is one of the most efficient methods for today's first principles calculations. Developed in 1975 by O.-K. Andersen (Max-Planck Institute for Solid State Research, Stuttgart, FRG), the method has

since been successfully used on many different systems: from pure metallic crystals to complex compounds and supercell models. In combination with the so-called atomic sphere approximation, this method is some two orders of magnitude faster than most of the traditional methods used earlier. In this work LMTO spin-polarized scalar-relativistic calculations were done on a uniform mesh of 112 points in the irreducible wedge of the hexagonal Brillouin zone. The exchange-correlation functional of von Barth and Hedin [20] and the frozen core approximation [21] were used [22] .

Results and Discussion.

Table 1 summarizes some of the results. As one can see, due to the strong electron hybridization, a significant charge-density redistribution among the atomic spheres was observed. Contrary to earlier molecular cluster calculations, a charge transfer [23] from the impurity towards the nearest Fe-atoms was discovered. This direction of charge transfer also contradicts what one would expect from simple electronegativity considerations: the impurities are more electronegative than Fe [24] . However, such a "reverse" charge transfer was previously found in PdB (see C. D. Gelatt et al, [23]), and Ni_3P [25] . Only H demonstrates an opposite behavior.

Like the case of the PdB compound, the observed redistribution of charge may be explained as a result of forming strong p-d hybrids with more localized bonding components centered around the Fe atoms. The loss of 1.61 electrons by the Wigner-Seitz (WS) sphere of P (Table 1) agrees well with the 1.45 electron charge transfer between the P and Ni WS spheres in Ni_3P [25] . Such a situation may be quite typical when the sp-d hybridization is involved.

Figure 2 shows the site-projected density of states (DOS) plots for GBs with B and C. On each plot the upper half,"positive" panels correspond to the host Fe atoms [(a) on the GB, Fe-1; or (b) across the GB,Fe-2], while the lower half panels are the DOSs for each impurity WS sphere (for the same impurity, in the plots (a) and (b), these half panels are identical).

Analysis of site-projected DOS plots sheds light on the most important features of electron behavior around a given atom. Whenever electronic states in the same energy range but with the different angular momentum (s-, p- or d-) are present both in an Fe atom and the impurity, they are "hybridized," contributing to covalent bonding. The more states that are involved in such a hybridization, the stronger the bonding. On the other hand,whenever the electronic states around a host atom (say, Fe) do not have their counterparts in the same energy range around the impurity atom, they are said to be "nonbonding": these states do not contribute to covalent bonding.

Figure 2 reveals a strong hybridization between the (mostly d-) electrons of Fe in the GB plane (a) on the one hand, and (mostly p-) electrons of B and C on the other hand. In a more relaxed GB supercell with C one would expect a more B-like picture, with a somewhat higher d-DOS in the energy range right above the p-electrons in the C-sphere.The hybridization is weaker for the Fe atoms across the GBs with B and C (Fig. 2, plots (b)), with a part of the impurities p- electrons occupying nonbonding states. However, one may claim that the Fe-B or Fe-C covalent bonding is still rather strong.
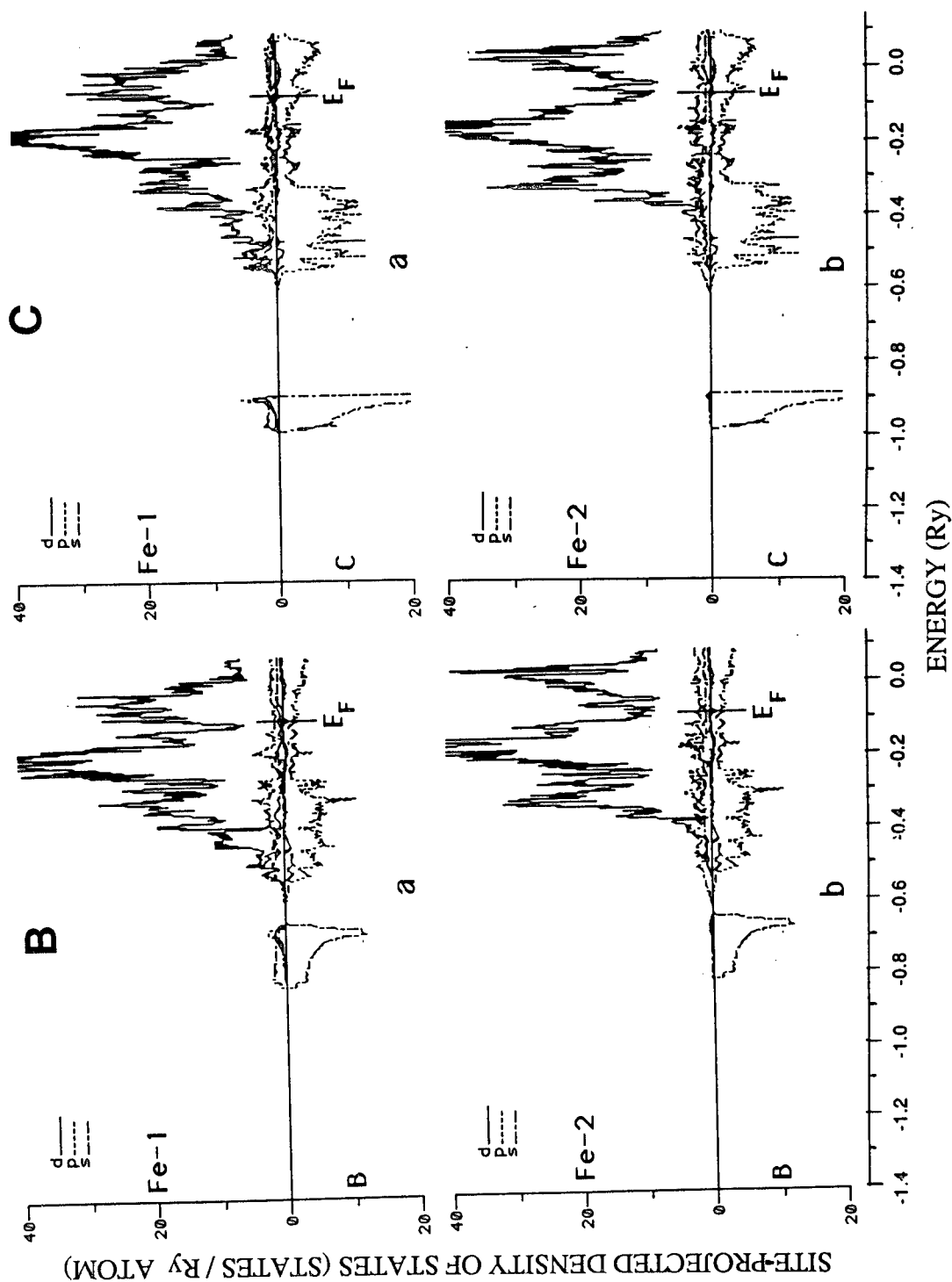
Fig. 2. The site-projected DOS plots for Fe-1 atoms (a) , Fe-2 atoms (b) and impurities for the model GB with C and B
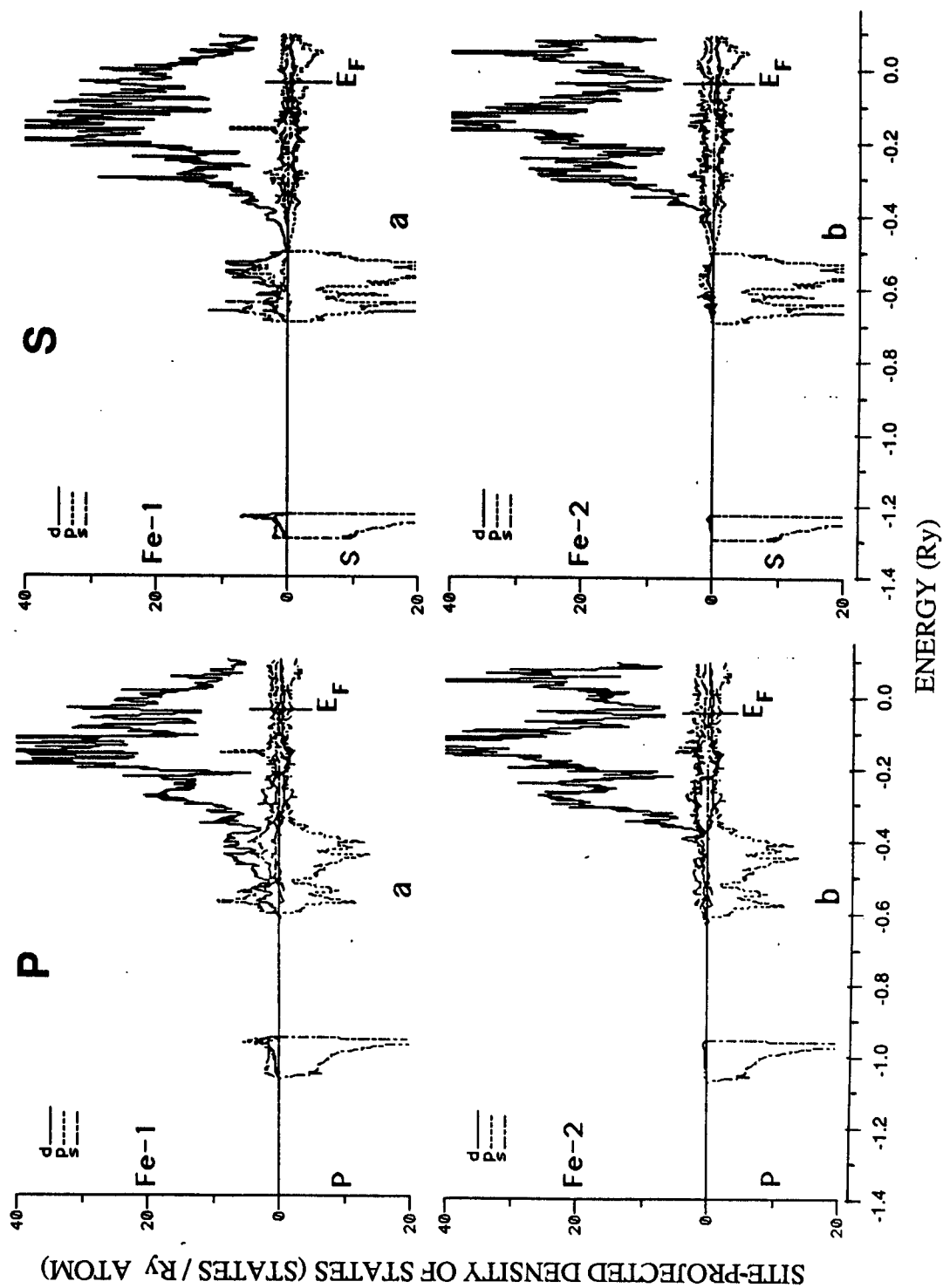
Fig. 3. The site-projected DOS plots for Fe-1 atoms (a) , Fe-2 atoms (b) and impurities for the model
GB with P and S

As one can see from Fig. 3 the bonding of Fe atoms on the GB with P and S (plots (a)) is also relatively strong, resulting from a significant p-d hybridization. The picture is profoundly different for the Fe-2 atoms across the GBs (b) with P and S. Here both Fe d-electrons and p-electrons of P and S have different energies, and the corresponding energy states are almost entirely nonbonding, resulting in a dramatic weakening of interatomic bonds across the GB.

Another interesting trend is observed for all the impurities. The so-called "impurity bands" --the energy states occupied mostly by the impurity s-electrons and separated from the main valence bands-- are shifted towards lower energies when one moves from B to S. The gap between the impurity band and the valence bands completely disappears for GB with H. This trend of course reflects the succession of s-levels in free impurity atoms.

It is seen from Fig. 2, plots (a),that the impurity s-electrons are also involved in bonding: there is a weak d-component in the Fe-1 DOS spectra in the energy range corresponding to the impurity bands -- far below the Fermi level. This component can not be explained as a mere "tail" from the nearby impurity. Another manifestation of the s-d hybridization is the decrease in the number of s-electrons (with respect to that in a free impurity atom), Table 1 , which "opens up" the closed 2s (B,C) or 3s (P,S) shells of the single impurity atoms for bonding.

Though the "many-particle" exchange-correlation contributions are essential for stabilizing a ferromagnetic system, the important "one particle" correlation of having a low DOS at the Fermi energy (E_F) for a stable situation is still manifested in Figs. 2 and 3 . For the Fe-1 atoms on GBs with both B or C as well as P or S, the site-projected DOSs at E_F fall in (or very close to) minima. For the Fe-2 atoms (across the GBs) the $DOS(E_F)$ still have narrow minima with B and C, but correspond to sharp peaks with P and S.

Figure 4 shows the analogous DOS plots for the GB with H. In case (a), there is some s-d hybridization (though much weaker than for other impurities). The hybridization almost disappears for the Fe-2 atoms across the GB. The plots for H illustrate the general trend in interatomic bonding if an impurity is placed into the center of the Fe trigonal prism. As was mentioned above, the true situation with the H atoms on a GB may be quite different..

Returning to Table 1, one can see that the magnetic moments of the Fe atoms on and across the GB are quite different. The Fe-1 atoms have magnetic moments lowered by 10-20% (B,C) and 40-50% (P,S) with respect to that of Fe in BCC phase ($2.2\mu_B$/atom); this is even greater if compared with Fe in a somewhat expanded BCC phase corresponding to an average atomic volume of the supercell (see Table 1). This result agrees with earlier preliminary calculations on a 22-atom GB supercell using the LMTO-Stoner approach [26]. On the other hand, the Fe-2 atoms have their moments significantly enhanced.

The suppression of the magnetic moments in Fe on the GB may be interpreted as a manifestation of the lesser filling of antibonding states, while, the enhancement of moments in atoms across the GB results in a higher antibonding state occupation. However, the net effect of this magnetic energy change may be quite the opposite, since the spin polarization (ferromagnetism) produces a negative (cohesive) contribution to the total energy, roughly proportional to the square of the magnetic moment. For example, the suppression of the magnetic moments of Fe atoms by a surface layer of Ni is almost entirely responsible for the increase in the effective surface energy [27] .
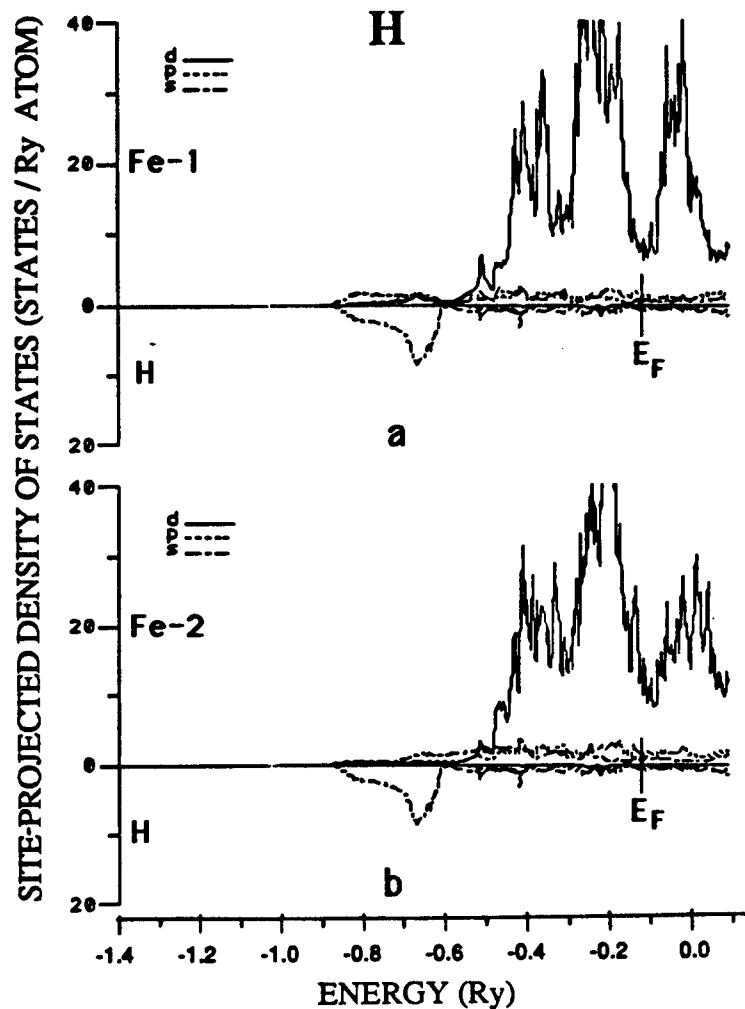
Fig. 4. The site projected DOS plots for Fe-1, Fe-2 and H atoms (GB with H)

Conclusions

It was found, that in a typical Fe GB environment, C and B on the one hand, and P , S and H on the other hand, demonstrate quite a different behavior. The former show strong bonding with the Fe atoms both in the GB plane and across the GB, while the latter dramatically weaken the bonding across the GB. These results should be important for understanding the decohesion mechanisms (which can produce embrittlement in Fe and Fe-base alloys) and could lead to improved Army steels..

They also suggest a way of improving intergranular cohesion. In order to "cancel out" the decohesive effect of an impurity (say, P or S) one should somehow change the GB environment so that the electron hybridization could be enhanced.This would be achieved by

"injecting" missing electronic states into the host DOSs, Fig.3,(b), to convert nonbonding states to bonding ones. The latter may be accomplished by suitable solute alloy additions.

At the present time, a more sophisticated GB model is being investigated. That research, together with the independent calculations on the change of the Fe surface energy with the addition of impurity atoms to the surface (Prof. Freeman's group, Northwestern University.) will enable one to build a more quantitative foundation for the prediction of the impurity embrittling effects. Such understanding will be an important input into theory supported design of new embrittlement resistant high strength steels and alloys for ballistic protection.

References

1. C. L. Briant and S.K. Banerji in Embrittlement of Engineering Alloys (ed. C. L. Briant and S.K. Banerji), Acad. Press, New York, 1983, p. 21; M. Guttmann and. D. McLean in *Interfacial Segregations* (ed. W.C. Johnson and J. M. Blakely), ASM, Metals Park, OH, 1979, p. 261
2. J. P. Stark and H. L. Marcus, Metall. Trans. A , **8A**, 1423 (1977); D. Y. Lee, E. V. Barrera, J. P. Stark and H. L. Marcus, Metall. Trans. A, **15A**, 1415 (1984)
3. C. L. Meyers, Jr., G. Y. Onoda, A. V. Levy, and R. J. Kotfila, Trans. Metall. Society of AIME, **233**, 720 (1965)
4. J. Komedo and C. J. McMahon,Jr, Metall. Trans. **12A**, 31 (1981)
5. M. P. Seah, J. Phys. F, **10**, 1043 (1980)
6. J. R. Rice and J.-Sh. Wang, Materials Science and Engineering, **A107**, 23, (1988)
7. A. R. Troiano, Trans. Am. Soc. Met. **52**, 54 (1960)
8. C. L. Briant and R. L. Messmer, Phil. Mag. B **42**, 569 (1980); R. L. Messmer and C. L. Briant, Acta Metall., **30**, 457 (1982)
9. C. M. Sayers, Phil. Mag. B **50**, 635 (1984)
10. M. E. Eberhart, K. H. Johnson and R. M. Latanision, Acta Metall. **32**, 955 (1984); M. E. Eberhart, R. M. Latanision and K. H. Johnson ,*ibid*, **33**, 1769 (1985); M. E. Eberhart and D. D. Vvedensky, Phys. Rev. Lett. **58**, 61 (1987)

11. G. S. Painter and F. W. Averill, Phys. Rev. Lett. **58**, 234 (1987)

12. M. Mori and Y. Ishida, in *Grain Boundary Structure and Related Phenomena,* Proc. of JIMIS-4 (1986). Supplement to Trans. of Japan Inst. of Met., p. 361

13. A. Ya. Belenk'ii and M. A. Fradkin, Poverkhnost'. Fizika, Chimiya, Mekhanika , **9**, 90 (1989)

14. M. Hashimoto, Y. Ishida, R. Yamamoto, M. Doyama and F. Fujiwara, J. Phys. F. **11**, L141 (1981); Surface Sci. **144**, 182 (1984); M. Hashimoto, Y. Ishida, S. Wakayama, R. Yamamoto, M. Doyama and F. Fujiwara, Acta Met. **32**, 13 (1984)

15. S. Crampin, D. D. Vvedensky, J. M. MacLaren and M. E. Eberhart, Phys. Rev. B **40**, 3413 (1989)

16. M. Hashimoto, Y. Ishida, R. Yamamoto, M. Doyama and T. Fujiwara, Scripta Met.,**16**, 267 (1982); M. Hashimoto, Y. Ishida, R. Yamamoto and M. Doyama , Acta Met. **32**, 1 (1984); Y. Ishida and M. Mori, Journal de Physique, Colloque C4, **46**, C4-465 (1985)

17. B. Aronsson and S. Rundquist, Acta Cryst. **15**, 878 (1962); S. Rundquist, Arkiv für Kemi, **20**, 67 (1962)

18. O. K. Andersen, O. Jepsen and D. Glötzel, in *Highlights of Condensed Matter Theory* (ed. F. Bassani, F. Fumi and M. P. Tosi) North Holland, New York, 1985; O. K. Andersen, in *Electronic Structure of Complex Systems,* (ed. P. Phariseau and W. M. Timmerman) , Plenum, New York, 1984, p. 11

19. H. L. Skriver, *The LMTO Method* , Springer, Berlin, 1984

20. U. von Barth and L. Hedin, J. Phys. C **5**, 1629 (1972)

21. U. von Barth and C. D. Gelatt, Jr., Phys. Rev. B **21**, 2222 (1980)

22. The frozen core electron configurations of (B,C) , (P,S) and Fe are respectively those of the inert He, Ne and Ar atoms.

23. The definition of charge transfer is quite controversial. See, e.g. discussion in C. D. Gelatt, Jr., A. R. Williams and V. L. Moruzzi, Phys. Rev. B **27**, 2005 (1983). If one uses the atomic sphere approximation, a "natural" way of defining the charge transfer is to compare the amounts of charge within different WS spheres, though, the ambiguity still remains with respect to assigning definite volumes to the WS spheres. One way of making charge transfer less ambiguous is to compare charge distributions of interest with that of a "standard" reference system. In this case the reference system can be chosen to be a "clean GB" (i.e. a supercell with an empty sphere substituted for the impurity) plus a single impurity layer (i.e. a supercell with the empty spheres substituted for all Fe-atoms). With respect to such a reference system it was still found that, for P, there is a negative charge transfer of 0.07 electrons.

24. The Pauling electronegativities of Fe, B, C, P, and S are respectively: 1.8, 2.0, 2.5, 2.1 and 2.5 (L. Pauling, *The Nature of the Chemical Bond,* Cornell University Press, Ithaca, N.Y., 1960, p. 93)

25. S. S. Jaswal, Phys. Rev. B **34**, 8937 (1986)

26. G. L. Krasko and G. B. Olson, Bull. Am. Phys. Soc. **34**, 772 (1989)

27. G. L. Krasko and G. B. Olson, in *Proceedings of the 34th Sagamore Army Materials Research Conference,* 1987 (in press)

Measurement of the Refractive Indices of
Fluoride Glasses using Digital Refractometry (U)

*Hon C. Kwan, CPT, SC
Steven S. Cotariu, CPT, IN
US Army Photonics Research Cell
United States Military Academy
West Point, New York 10996-5000

During the past twenty years requirements have been increasing for new optical glasses for use in military optical applications, e.g., fiber optics, optical windows, lens systems, and detectors. Many of these glasses are needed for use in the infrared region of the optical spectrum. The new fluoride glasses being developed show great potential for use in these infrared systems. These glasses are expected to find significant application in military and commercial fiber optic and optical lens systems.[1]

A number of classical methods have been developed for measuring the refractive index of glass, such as the minimum deviation prism method and the Brewster angle null method. These methods, however, are time consuming, expensive, and difficult to implement, especially in the infrared region. They also require extensive preparation of a glass sample which may have to be cut into a relatively large prism, then ground and polished. Most fluoride glasses, e.g., ZBLAN-Er-Tb, are extremely fragile and are not tolerant of the stress encountered during preparation for conventional refractive index measurement methods. A new method was needed to determine the refractive index of fluoride glasses in a faster, simpler, less expensive, and non-destructive manner.

We have applied Digital Refractometry[2] to the measurement of the refractive index of four new fluoride glasses. Digital Refractometry is based on a change in polarization state of light reflected from an optically flat surface at an incident angle of $45°$. At this angle, Fresnel's

reflection equations simplify and lead to a new equation to determine the refractive index, n. We begin our paper with a brief description of Digital Refractometry and discuss its advantages over other methods for measuring the refractive index. We then describe the measurement of the refractive index of four new fluoride glasses for visible and near infrared wavelengths. These measurements are then compared to the refractive indices reported by other researchers using the minimum deviation prism method.[3,4]

THEORY

Digital Refractometry is a method for measuring the refractive index of materials using readily available optical and electronic components. The theory of Digital Refractometry rests on a simplification in Fresnel's equations for the reflected intensity of polarized light at an incident angle of $45°$ and the availability of highly accurate digital voltmeters.

To describe Digital Refractometry, we consider the reflection of light from an optical surface as shown below:
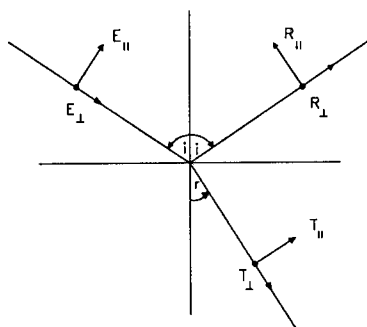
Figure 1. Polarization states of an incident, reflected, and refracted beam.

The incident ray consists of polarized perpendicular and parallel components E_s and E_p, respectively. The beam is reflected from the surface with perpendicular and parallel components R_s and R_p, respectively. Fresnel's equations for the polarized reflected components are:[5]

$$R_s = -\left[\frac{\sin(i-r)}{\sin(i+r)}\right] * E_s \; , \tag{1a}$$

$$R_p = \left[\frac{\tan(i-r)}{\tan(i+r)}\right] * E_p \; . \tag{1b}$$

The two special cases for these equations most often discussed are for normal incidence ($i = 0$) and the Brewster angle ($i = i_B$).

It is simpler to describe polarization in the intensity domain by transforming Fresnel's equations to a 4 X 4 Mueller polarization matrix.[6] Upon doing this one finds,

$$\begin{pmatrix} s_{0r} \\ s_{1r} \\ s_{2r} \\ s_{3r} \end{pmatrix} = \frac{\tan^2\alpha_-}{2\sin^2\alpha_+} \begin{pmatrix} \cos^2\alpha_- + \cos^2\alpha_+ & \cos^2\alpha_- - \cos^2\alpha_+ & 0 & 0 \\ \cos^2\alpha_- - \cos^2\alpha_+ & \cos^2\alpha_- + \cos^2\alpha_+ & 0 & 0 \\ 0 & 0 & -2\cos\alpha_-\cos\alpha_+ & 0 \\ 0 & 0 & 0 & -2\cos\alpha_-\cos\alpha_+ \end{pmatrix} \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} \; . \tag{2}$$

where $\alpha_- = i-r$ and $\alpha_+ = i+r$. The column matrices are the Stokes vectors, S_r and S, for the reflected and incident fields, respectively. At an incident angle of $45°$, equation (2) reduces to a Mueller matrix of the form,

$$M_r = \frac{1-\sin 2r}{(1+\sin 2r)^2} \begin{pmatrix} 1 & \sin 2r & 0 & 0 \\ \sin 2r & 1 & 0 & 0 \\ 0 & 0 & -\cos 2r & 0 \\ 0 & 0 & 0 & -\cos 2r \end{pmatrix} \; . \tag{3}$$

This simplified Mueller matrix and Snell's Law are used to calculate the refractive index of the material. To derive an equation for the refractive index, assume the optical surface is irradiated by a linearly polarized beam with its plane of polarization at $45°$. By using a polarizer as an analyzer of the reflected beam, two unique equations are derived from equation (3) for the reflected beam. The analyzer is first rotated to the s-position and the intensity is found to be

$$I_s = I_o \left[\frac{1 - \sin 2r}{1 + \sin 2r} \right]. \tag{4}$$

Next, the analyzer is rotated to the p-position and the intensity is then found to be

$$I_p = I_o \left[\frac{(1-\sin 2r)^2}{(1+\sin 2r)^2} \right]. \tag{5}$$

The value of the intensity of the first equation (4) is larger than the second (5), so I_s is called I_{max} and I_p is called I_{min}. Taking the ratio of I_{max} to I_{min} and solving for $\sin 2r$ gives

$$\sin 2r = \frac{(I_{max} - I_{min})}{(I_{max} + I_{min})}. \tag{6}$$

This equation can be solved for the refractive index by substituting Snell's Law for r. After a considerable amount of algebra, this leads to the following equation,

$$n = \frac{\sqrt{R+1}}{\sqrt{R-1}}, \tag{7}$$

where $R = I_{max}/I_{min}$.[7]

In Digital Refractometry, it is only necessary to measure the intensities of the s and p polarization components of the reflected beam to determine the refractive index. This can be accomplished using equipment normally available in most optical laboratories. The advantages of Digital Refractometry over conventional methods are many. Among these are measurement speed, minimal preparation of the sample, wide spectral application, and the relatively simple automation of the measurement.

The accuracy of the measurement depends upon several factors: the angular resolution of the polarizer mounts, the quality of the material surface, the quality of the polarizers, the accuracy of the angle of the incident and reflected beam, the linearity of the detector, the significant digits on the digital voltmeter, the stability of the optical source, and the overall presence of noise in the system. We emphasize that to apply Digital Refractometry, only a small region of an optical surface must be prepared.

MEASUREMENTS

The laboratory implementation of Digital Refractometry is shown in Figure 2. A calcite linear polarizer with its plane of polarization at $45°$ was placed between the optical source and the sample. A second analyzing calcite polarizer in a rotating mount allowed us to measure the perpendicular and parallel components of the reflected light from the sample.

Initially, a Spectra Physics Model 117A Intensity Stabilized 632.8 nm Helium Neon laser was used as the optical source. The optical and mechanical alignment was checked by measuring the refractive index of a sample of laser quality Schott BK7 glass. The sample was positioned to within 1 arc-second of $45°$ to the incident beam using a Newport Model 610 Ultra-Resolution Mirror Mount, a Gaertner 551 autocollimating telescope, and interferometric techniques. A UDT DP10 silicon photodiode was used to detect the reflected beam. The DP10 signal output was connected directly into the transformer coupled input of an EG&G 5316 preamp mounted in a EG&G 5301 lock in amplifier. The analog output voltage of the lock in amplifier was measured by a 8 1/2 digit HP 3458A digital multimeter under GPIB control of an AT compatible personal computer.
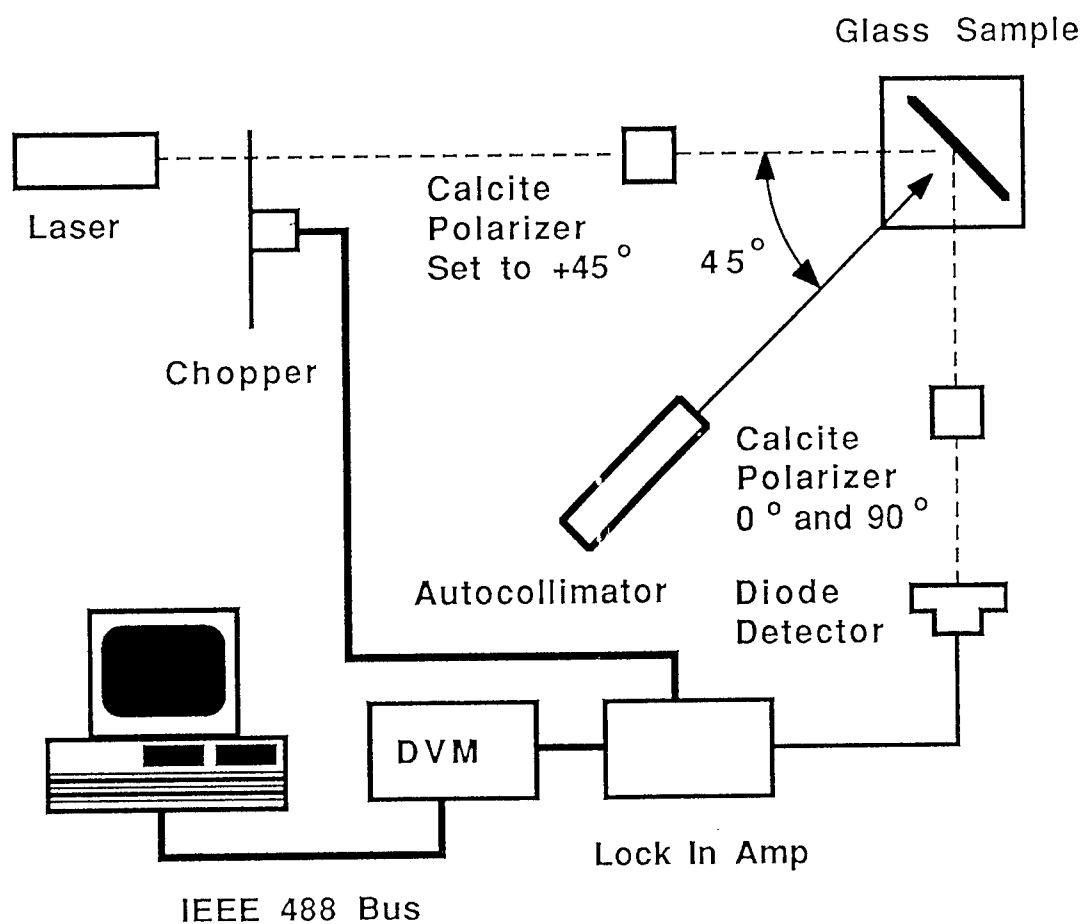
Figure 2. Experimental Apparatus
for Digital Refractometry

By careful alignment of components, reduction of noise to the lowest possible levels, and by averaging the detected signal for both polarizations over hundreds of data readings, a measurement of 1.51452 ± 0.00022 was obtained for BK7 glass, as compared to 1.51509, as given by the Schott Glass catalog.[8]

The ZBLAN samples were 1 inch disks and presented no problems for alignment or measurement. The CLAP glass severely tested the flexibility and adaptability of the technique. The CLAP samples were extremely small and irregularly shaped. They ranged from small flat disks 5 mm across, to square rods 2 mm square on the ends and 1 cm long, to a small blob about 8 mm in diameter. Two samples were optically polished on the most convenient side, then cemented to a 2 inch diameter glass disk. This assembly was then inserted into the Newport Model 610 Ultra-Resolution mount. The first set of measurements were made using the 632.8 nm Helium Neon laser. The next sets of measurements were made using thermoelectrically cooled Melles Griot semiconductor laser diodes, models 06 DLL 207 and 06 DLL 507 emitting at 786 nm and 824 nm, respectively.

DATA

The calculation of the refractive index at a particular wavelength required measurement of I_{max} and I_{min}. A computer program was written using Lotus Measure and Lotus 123 macro programming commands to automate the data collection. The program automatically acquired the data and entered it into a spreadsheet for analysis and future reference. At the beginning of a data collection run, the analyzing polarizer was manually set to the perpendicular (I_{max}) position. The computer controlled multimeter then measured the voltage for I_{max} 100 times at 1 second intervals. The data acquisition was halted temporarily to allow manual rotation of the analyzing polarizer to the parallel (I_{min}) position. The program was resumed, collecting 100 measurements of I_{min}. The refractive index was calculated from the average value of I_{max} and the average value of I_{min} over the 100 values. A single data

run took about 10 minutes to complete. Five data runs were
made for each type of glass for each different wavelength.
The final value of the refractive index and the standard
deviation were calculated by averaging the values of the
five data runs. The first data runs for each wavelength
were always for BK7 glass. This allowed calibration of the
experiment to a known value before taking data for the
fluoride glasses. The final results of the five data runs
for each wavelength and type of glass are presented in
tabular form. The measured values are the averaged index

of refraction ± the averaged standard deviation obtained
using Digital Refractometry. The second values are the
index of refraction calculated from dispersion coefficients
extracted from minimum prism deviation prism measurements,
as reported by other researchers.

Table 1.

Schott BK7 Glass

Wavelength (nm)	Digital Refractometry	Min Dev Prism	% Difference
632.8	1.51452 ± 0.00022	1.51509	-0.03762
786	1.50916 ± 0.00008	1.51105	-0.12508
824	1.51227 ± 0.00012	1.51031	+0.12977

Table 2.

ZBLAN Fluoride Glass

Wavelength (nm)	Digital Refractometry[*]	Min Dev Prism[**]	% Difference
632.8	1.50601 ± 0.00027	1.49657	+0.63078
786	1.50014 ± 0.00013	1.49365	+0.43451
824	1.50573 ± 0.00008	1.49313	+0.84386

Mole Percent Compositions:

[*] ZrF_4 (53) - BaF_2 (20) - LaF_3 (4) - AlF_3 (3) - NaF (20)

[**] ZrF_4 (52) - BaF_2 (19) - LaF_3 (5) - AlF_3 (4) - NaF (20)

Table 3.

ZBLAN-Er-Tb Fluoride Glass

Wavelength (nm)	Digital Refractometry[*]	Min Dev Prism[**]	% Difference
632.8	1.50165 ± 0.00033	Not Available	-
786	1.50987 ± 0.00008	Not Available	-
824	1.50016 ± 0.00013	Not Available	-

[*] ZrF_4 (53) – BaF_2 (20) – LaF_3 (4) – AlF_3 (3) – NaF (20) with additions of ErF_3 and Tb_3

[**] ZBLAN-Er-Tb too fragile for Minimum Deviation Prism Methods

Table 4.

CLAP Fluoride Glass

Wavelength (nm)	Digital Refractometry[*]	Min Dev Prism Method[**]	% Difference
632.8	1.59119 ± 0.00035	1.60894	-1.10321
786	1.59845 ± 0.00010	1.60208	-0.22658
824	1.58501 ± 0.00036	1.60090	-0.99257

[*] CdO (13) – CdF_2 (9) – LiF (10) – AlF_3 (34.5) – PbF_2 (33.5)

[**] Same composition as measured sample.

Table 5.

CLAP-Ky-La Fluoride Glass

Wavelength (nm)	Digital Refractometry*	Min Dev Prism**	% Difference
632.8	1.59220 ± 0.00019	1.59275	-0.03453
786	1.59065 ± 0.00010	1.58625	+0.27738
824	1.57745 ± 0.00018	1.58518	-0.48764

* CdO (12) – CdF_2 (12) – LiF (6) – AlF_3 (30) – PbF_2 (30) – KF (4) – YF (4) – LaF_3 (2)

** Same composition as measured sample.

ANALYSIS

Measurement of the fluoride glass was only accomplished after an alignment check of the experiment. Alignment was accomplished by inserting BK7 into the sample holder, aligning the system, and measuring the refractive index. This measurement could then be compared to the known values for BK7 listed in the Schott Glass Catalog. Refractive index values not directly available from the catalog were determined using the dispersion equation coefficients. If the measurement was within 1% of the known value, the experiment proceeded to the fluoride glasses.

Manual rotation of the analyzing polarizer mount appeared to be the single largest source of error in the experiment. Between data runs the measured index of refraction varied by as much 0.2 %, but the variation within a run was typically an order of magnitude less. We determined that the angular resolution of the analyzing polarizer mount could account for nearly all of the variation between runs. The repeatable resolution of the mount was

0.05°. Calculating the index of refraction using a $\Delta\theta = 0.05°$ caused the index to vary by 0.002 (0.17 %).

The ZBLAN and CLAP glass samples were polished to approximately a 60/40 scratch/dig surface. The BK7 glass was polished to a laser grade surface. This uncertainty in the surface preparation of the samples may explain why the percent difference for BK7 was generally lower than the percent differences for ZBLAN and CLAP glasses.

There was also some uncertainty with the exact compositions of the glass samples. The ZBLAN glass sample had a slightly different mole percent composition than the ZBLAN glass measured using the minimum deviation prism method. CLAP glass is currently made one small batch at a time. There may be slight variations in composition between batches.

We believe that three specific actions can reduce the difference between the measured values and the minimum deviation prism method values. First, replace the manually rotated analyzing polarizer mount (0.05° resolution) with a computer controlled motorized rotary stage (0.001° resolution). This would reduce the variance between data runs by minimizing human error. Second, polish the sample materials to the absolute best surface condition possible. Third, cool the detector and use ultra low noise electronics.

CONCLUSIONS

The refractive indices of several new fluoride glasses, called ZBLAN and CLAP glasses, were measured for three different wavelengths using Digital Refractometry. The results were in good agreement with values obtained from previously reported minimum deviation prism measurements. Digital Refractometry proved to be a versatile and useful method for rapidly determining the refractive index of fluoride glasses.

ACKNOWLEDGEMENTS

We wish to thank Dr. Edward Collett, the developer of Digital Refractometry, and Mr. Thom Brown of Measurements Concepts in Colts Neck, New Jersey for their assistance in conducting the research. The CLAP glass samples were

REFERENCES

1. G. H. Sigel, Jr., "Fluoride Fiber Progress in the USA", Materials Science Forum, Vol. 32-33, 283 (1988).

2. E. Collett, "Digital Refractometry", Opt. Commun. Vol. 63, 217 (1987).

3. S. Mitachi and P. Tick, "Dispersion Characteristics of CLAP Glass Systems", Proc. SPIE, San Diego, August 18 & 19, 1988.

4. L. Wetenkamp, T. Westendorf, G. West and A. Kober, "The Effects of Small Composition Changes on the Refractive Index and Material Dispersion in ZBLAN Heavy-Metal Fluoride Glass", Materials Science Forum, Vol. 32-33, 471 (1988).

5. M. Born and E. Wolf, "Principles of Optics", Pergamon Press, New York, (1975).

6. E. Collett, "Mueller-Stokes Matrix Formulation of Fresnel's Equations", American Journal of Physics, Vol. 39, 517 (1971).

7. A. Ballato, unpublished equation.

8. Schott Catalog-Optical Glass, Schott Optical Glass, Inc., York Ave., Duryea, PA 18642, USA.

Experimental Polarization Dependence of Bragg Diffraction
Using Surface Acoustic Waves in Lithium Niobate (U)

*H. C. Kwan, CPT, SC; C. DeCusatis, Dr.; D. M. Litynski, COL
US Army Photonics Research Cell
United States Military Academy
West Point, New York 10996-5000

Introduction

Many signal processing applications require the ability to analyze enormous amounts of information at very high speed. Optical information processing techniques are capable of satisfying these requirements; in particular, the diffraction of light through acousto-optic (AO) interaction has been studied for many applications. For example, AO devices have been developed for fundamental operations such as convolution and correlation.[1-4] Matched filters [4-6] and real time pattern recognition using phase-only filters may both be realized using AO processors. More generally, multiple transform AO devices can realize complex operations such as generating cross-ambiguity functions, Wigner-Ville distributions, and Mellin transforms.[1,4]

More advanced device concepts include optical channelizers, which would utilize integrated electronic detector arrays and Bragg cell technology to measure frequency and amplitude over a large dynamic range. For interference suppression, optical null steering processors using AO devices adaptively suppress wideband interference from multiple jammers.[1] These techniques may be combined with optical control systems for beam forming and steering in phased array antennas.[1] Finally, the growing field of optical computing has generated increased interest in the optical realization of neural networks and related forms of associative memory processing systems. [7]

Given this motivation, surface acoustic wave (SAW) devices have become increasingly important. These devices exhibit low driving power compared to bulk wave devices, and the design of interdigital transducers (IDTs) allows for versatile signal processing operations. For many applications, their compact, rugged format makes them well suited for integrated optics implementations. In particular, piezoelectric crystals such as lithium niobate ($LiNbO_3$) are well suited to the fabrication of SAW devices by basic photolithography procedures. Direct application of a drive voltage to the crystal IDT will excite surface waves. However, such materials have anisotropic crystal

structures; while this may enhance their performance in optical signal processors, the behavior of these media is not fully understood. For example, recent experiments [8] using one of the most common cuts of LiNbO$_3$ (y-cut, z-propagating) have confirmed that incident light polarization has a significant effect on anisotropic AO interaction. Effects such as different angular bandwidths, diffraction efficiencies, and a degeneracy in the Bragg diffraction condition were observed for different incident polarizations. While some of these results were later described qualitatively,[9] the effects lack a clear theoretical context and a full quantitative description.

We have conducted experiments at the Photonics Research Cell of the U.S. Military Academy, West Point, to independently confirm the effects reported for SAW interaction in lithium niobate. These experiments were then extended to higher acoustic frequencies; a coupled mode theory which partially explains the experimental results has been developed by collaborators at Rensselaer Polytechnic Institute, Troy, NY.[10] Our work also utilizes sophisticated computer controlled, automated data acquisition techniques for accuracy and repeatability not previously reported. We will first describe the theory behind SAW acousto-optic interaction in anisotropic media, which explains why the diffracted light intensity depends on incident light polarization. The SAW is modeled as a superposition of longitudinal and transverse bulk wave components, and each component may be analyzed separately using coupled mode theory. A propagating SAW is characterized by the strain present along a given direction in the crystal, which perturbs the local refractive index. Only first order index perturbations will be considered. In the following analysis, we assume a y-cut, z-propagating crystal of LiNbO$_3$ to be the anisotropic media, although the principles of the analysis are valid for any media or crystal cut. Although the theory also yields results for Raman-Nath diffraction, we will concentrate on effects in the Bragg regime because of its many signal processing applications. Both possible polarizations of incident light will be considered. A computer simulation of our model yields results which are in good agreement with recent experimental data.

SAW Interaction Theory

The crystal geometry and coordinate system assumed throughout our calculations are shown in Fig. 1. The behavior of anisotropic crystals may be characterized by the index ellipsoid, which takes the following form in a uniaxial crystal:

$$\frac{x^2}{n_o^2} + \frac{y^2}{n_o^2} + \frac{z^2}{n_e^2} = 1 \tag{1}$$

where n_o and n_e are the ordinary and extraordinary principle refractive indices. The magnitude of the applied acoustic strain may be related to refractive index perturbations through the elasto-optic coefficient matrix, as follows:

$$[\triangle(1/n^2)] = [p_{ij}] \quad [s_j]$$
$$(6 \times 1) \qquad (6 \times 6)(6 \times 1) \tag{2}$$

where s_j is the magnitude of the applied strain along the j direction, and p_{ij} are the elasto-optic coefficients for a given material.[11] We will consider the refractive index perturbations due to each component of the SAW separately. In the case of piezo-electric crystals, the applied strain produces an electric field, which further perturbs the refractive index through the electro-optic effect.[12] Although we have included this effect in our model, we will continue to describe the AO interaction in terms of the elasto-optic coefficients p_{ij}. The dependence of refractive index on electric field will be understood in the following development, to emphasize the principles of our analysis.

Consider first the longitudinal component of the SAW, that is, the perturbations which occur only along the direction of SAW propagation. For this case, the only nonzero strain present is s_{zz} or s_3, and the resulting index ellipsoid perturbation is given by:

$$\triangle(\frac{1}{n^2}) = x^2 p_{13}s_3 + y^2 p_{23}s_3 + z^2 p_{33}s_3 \tag{3}$$

where n is the index of refraction. Adding this to the unperturbed ellipsoid, and using the fact that $p_{13} = p_{23}$ for LiNbO$_3$, yields

$$\frac{x^2}{(n_o')^2} + \frac{y^2}{(n_o')^2} + \frac{z^2}{(n_e')^2} = 1 \tag{4}$$

where we express the refractive indices by using a binomial expansion and neglecting higher order terms,

$$n_o' \cong n_o - \frac{1}{2}p_{13}n_o^3 s_3 \qquad n_e' \cong n_e - \frac{1}{2}p_{33}n_e^3 s_3 \tag{5}$$

Note that the above expressions contain two terms, the first representing the unperturbed index of the crystal and the second representing an index perturbation. Since the applied strain is sinusoidal, the index expressions are of the form

$$n_o' \cong n_o - \triangle n_o \sin(\omega_s t - k_s z) \tag{6}$$

where k_s = wave vector of the acoustic signal; $\omega_s = 2\pi f_s$, f_s = acoustic frequency; and

$$\triangle n_o = \frac{1}{2}p_{13}n_o^3 s_3 \qquad \triangle n_e = \frac{1}{2}p_{33}n_e^3 s_3 0. \tag{7}$$

If the refractive index perturbations can be expressed in this form, the resulting diffraction of incident light may be analyzed using coupled mode theory.[4] In the transverse case, perturbations in the direction normal to SAW propagation (in this case, the y-direction) are considered. The only nonzero strain component is now s_{yz}, or s_4, and the resulting perturbation of the index ellipsoid is given by

$$\triangle\left(\frac{1}{n^2}\right) = x^2 p_{14} s_4 - y^2 p_{14} s_4 + yz p_{44} s_4 \tag{8}$$

since $p_{14} = -p_{24}$ for LiNbO$_3$. In order to remove the yz cross term, we perform a similarity transformation[10] which involves rotating the coordinate system about the optic axis through an angle θ, defined by

$$\tan 2\theta = \frac{-2p_{44}s_40}{\frac{1}{n_o^2} - p_{14}s_40 - \frac{1}{n_e^2}} \tag{9}$$

Using a binomial expansion for the refractive index in the new coordinate system, neglecting higher order terms, and using the small angle approximation, the indices in the new coordinate system become

$$n_y' \cong n_y \cong n_o - \frac{1}{2}p_{14}n_o^3 s_4 \sin(\omega_s t - k_s z)$$
$$n_z' \cong n_e \tag{10}$$

Now that both longitudinal and transverse components of the SAW have been expressed in the form of a constant term plus a sinusoidal perturbation, the diffraction of incident light may be analyzed using coupled mode theory. The underlying theory has been reported elsewhere;[1,4] we only present the results of applying this theory to anisotropic AO interaction.

First, consider the longitudinal case, where the refractive index variation in the y and z directions is given by eq. (5). The incident light beam will be in the xz plane, at an angle ϕ with respect to the x-axis. Its electric field vector may be expressed as

$$\vec{E}_{\text{in}} = E_o e^{j(\omega_\ell t - k_\ell(x\cos\phi + z\sin\phi))} \tag{11}$$

where E_o is the incident light amplitude, k_ℓ is the wave vector, and $\omega_\ell = 2\pi f_\ell$ (f_ℓ being the light frequency). The diffracted light output will be of the form

$$\vec{E}_{\text{out}} = E_{\text{in}} \sum_{p=-\infty}^{\infty} A_p(L)e^{jp(\omega_s t - k_s z)} \tag{12}$$

where $A_p(L)$ represents the diffracted light amplitude of the pth order as a function of the interaction length, L.

For the case of Bragg diffraction, there is only a single diffracted order and the undiffracted light;

$$\vec{E}_{\text{out}} = \vec{E}_{\text{in}} [A_o(L) + A_1(L)e^{j(\omega_s t - k_s z)}] \tag{13}$$

In the transverse case, it was necessary to rotate the coordinate system within the crystal, as defined by eq. (9). The electric field vector of the incident light must now be resolved into two components along the crystal's coordinate axes in order to analyze the AO diffraction effect. Once the diffracted light has passed through the crystal, the inverse coordinate transformation is applied in order to express the output light distribution in the same coordinate system as the input light. This process will, in general, produce a change in polarization between the input and output light.

Consider, then, a y-polarized beam of incident light, E_{in}, which may be expressed in the form of eq. (11). By applying the coordinate transformation of eq. (9), the electric field vector of the light passing through the crystal is represented in (x',y',z') coordinates as

$$\vec{E} = (\cos\theta\hat{y}' + \sin\theta\,\hat{z}')E_{\text{in}} \tag{14}$$

where the notation $\hat{y}'$ indicates a unit vector in the y' - direction. The indices of refraction in the $\hat{y}'$ and $\hat{z}'$ directions are given by eq. (10). Note that since there is no refractive index perturbation in the z' direction, only the y'-polarized light undergoes diffraction.

However, since $n_{y'} > n_{z'}$ the z' polarized beam travels faster through the crystal interaction length, L. In other words, the optical path lengths of the y' and z' polarized beams are different, and the y' beam experiences a phase shift relative to the z' beam. This phase shift can be expressed as

$$e^{j\delta}, \delta = kL(n_{y'} - n_{z'}) \tag{15}$$

so that the light leaving the crystal is of the form

$$\vec{E}'_{\text{out}} = E_{y'}e^{j\delta}\hat{y}' + E_{z'}\hat{z}' \tag{16}$$

KWAN, DECUSATIS & LITYNSKI

Applying the inverse coordinate transformation, the observed ouput light in the Bragg regime becomes

$$\vec{E}_{\text{out}} = E_{\text{in}}[\hat{y} \; cos^2\theta[A_0 + A_1 e^{j(\omega_s t - k_s z)}]e^{j\delta}$$
$$+ \hat{y} \; sin^2\theta \qquad + \hat{z} \; cos\theta \; sin\theta$$
$$- \hat{z} \; cos\theta \; sin\theta \; [A_0 + A_1 e^{j(\omega_s t - k_s z)}]e^{j\delta}] \qquad (17)$$

Recall that this form is valid for y-polarized incident light. If the incident light is of the form of eq. (11) but is z-polarized, then we must once again resolve the light into its y′ and z′ components. For this case, the electric field of the light inside the crystal becomes

$$\vec{E} = (\sin \hat{y}' + \cos\theta \; \hat{z}')E_{\text{in}} \qquad (18)$$

with $n_{y'}$ and $n_{z'}$ still given by eq. (10). The results are similar in form to those obtained previously; the output light in the Bragg regime becomes

$$\vec{E}_{\text{out}} = E_{\text{in}}[\hat{y} \; cos\theta \; sin\theta \; [A_0 + A_1 e^{j(\omega_s t - k_s z)}]e^{j\delta}$$
$$+ \hat{y} \; sin\theta \; cos\theta \; + \hat{z} \; cos^2\theta$$
$$- \hat{z} \; sin^2\theta[A_0 + A_1 e^{j(\omega_s t - k_s z)}]e^{j\delta}] \qquad (19)$$

A surface acoustic wave (SAW) can be expressed as a superposition of longitudinal and transverse bulk wave components. Thus we may predict AO-SAW diffraction by superimposing the results obtained previously.

We must first describe the diffracted light intensity for $\phi \cong \phi_B$. In other words, we would like to describe the intensity distribution in the single diffracted order as the incident beam moves slightly away from the Bragg angle. In order to do this, we must return to coupled mode theory. The diffracted light output is assumed to be of the form of eq. (12); substituting this into the wave equation yields the following differential equation for the diffracted light amplitudes:

$$\frac{\partial A_p}{\partial z} - \frac{\pi \Delta n_o}{\lambda_o}[A_{p-1} - A_{p+1}] = \frac{jpQ}{2\ell}(p - 2\alpha)A_p \qquad (20a)$$

where

$$Q = 2\pi L\lambda_o/\lambda_s^2 n_o \qquad (20b)$$

$$\alpha = \left(\frac{n_o \lambda_s}{\lambda_o}\right) \sin\theta$$

and the second derivative $\partial^2 A_p/\partial z^2$ has been neglected, along with the term containing $\Delta(n_o)^2$.

For $Q \gg 1$ and $\alpha = 1/2$, eq. (20) has solutions in the Bragg regime. In particular, a set of coupled difference differential equations are formed which can be solved for A_0 and A_1.[1]

Solving these equations for the intensity distribution of the dc and diffracted light yields

$$| A_0 |^2 = I_0 = 1 - I_1 \tag{21a}$$

$$| A_1 |^2 = \left[\frac{\pi \Delta n_o L}{\lambda \sigma} \sin\sigma\right]^2 \tag{21b}$$

where

$$\sigma = \frac{1}{2}\left[(Q(1-2\alpha))^2 + \left[\frac{2\pi \Delta n_o L}{\lambda}\right]^2\right]^{1/2} \tag{21c}$$

These expressions will be used to calculate the intensity distribution of the diffracted light as a function of angle.

Using the above expressions, we have performed a computer simulation of our model. In order to generate these results, the ratio of longitudinal to transverse strain in the crystal was required, as well as the ratios of the induced electric field components in piezoelectric media. Expressions for the deformation caused by propagating surface waves have been reported;[13] we have used these results to estimate the strain ratios. This calculation was performed at the crystal surface, neglecting the decay of SAW components in the bulk crystal. This approximation is justified, since the surface wave decays almost completely within 1-2 wavelengths of the crystal surface. The variation of electric field components produced by interdigital transducers has also been reported,[12] and was used to calculate the effects of induced electro-optic effects.

Since the anisotropy of the material is significant, we must also redefine the Bragg angle[14] as follows:

$$\sin\theta_i = \lambda/2n_i\lambda_s[1 + \lambda_s^2/\lambda^2(n_i^2 - n_d^2)]$$
$$\sin\theta_d = \lambda/2n_d\lambda_s[1 - \lambda_s^2/\lambda^2(n_i^2 - n_d^2)] \tag{22}$$

where n_i and n_d are the refractive indices for the incident and diffracted light, respectively, λ is the wavelength of light, and λ_s is the acoustic wavelength.

Using these expressions, computer simulations of SAW interaction in lithium niobate at different acoustic frequencies and incident light polarizations were conducted by collaborators at Rensselaer Polytechnic Institute; typical results are shown in Fig. 2. The y-polarized light exhibits a sinc² diffraction efficiency near the Bragg angle, as expected from classical theory.[1,2] However, the z-polarized light exhibits an intensity dip near the Bragg angle; this is a new effect in anisotropic SAW diffraction, first observed in 1985.[8] We believe this effect results from the interaction of competing terms in eq. (21b). In lithium niobate, the refractive index for y-polarized light remains constant for all angles of incidence, and diffraction efficiency follows the sinc² form of eq. (21b). For z-polarized light, the refractive index increases with incident angle, ϕ, according to the index ellipsoid surface

$$\frac{1}{n^2} = \frac{\sin^2 \phi}{n_o^2} + \frac{\cos^2 \phi}{n_e^2} \tag{23}$$

Although the change in n is small, the change in Δn at large θ can be significant because Δn varies as n^3 (see eq. (7) or (10)). Initially, θ is small with respect to the Bragg angle, so diffraction efficiency follows the sinc² form of eq. (21b), increasing to the first peak observed in our simulations and then beginning to decline. Near the Bragg angle, however, the increase in Δn with θ becomes significant. Diffraction efficiency is also proportional to Δn; the increasing Δn overcomes the declining sinc² tail of eq. (21b), forcing diffraction efficiency to increase again and creating the second peak. As θ continues to increase, eventually the sinc² function dominates once again, forcing the diffraction efficiency to decrease from the second peak to zero. Additionally, our simulations indicate that the relative size of the dip depends on the magnitude of the applied acoustic strain; this is reasonable, since Δn is also a function of acoustic strain by eq. (7) and (10). Note that this effect also causes the z-polarized light to exhibit a larger angular bandwidth than y-polarized light; the different diffraction efficiencies observed for the two polarizations result from the relative magnitude of the elasto-optic coefficients for y and z-polarized light interactions.

Experimental Results

Recent experiments conducted at Harry Diamond Laboratories[9] reported the effects of polarization on Bragg diffraction efficiency in y-cut, z-propagating LiNbO₃. For acoustic frequencies in the range 90-350 MHz, an increased angular bandwidth and midband intensity dip were observed for z-polarized light, but not for y-polarized light. This agrees with our theory. One of the SAW devices used in the HDL experiments was available for our studies. The crystal geometry and coordinate system were the

same as in Fig. 1. To achieve a broad acoustic bandwidth, the SAW device employed the split finger transducer design illustrated in Fig. 3. The center frequency was 180 MHz, with a 60 MHz bandwidth; the device also functioned at frequencies near the 90 MHz subharmonic. This enabled us to confirm the anisotropic diffraction effects at frequencies where diffraction efficiency was higher and where the midband dip phenomena became discernible. Using recently installed facilities at the West Point Photonics Research Cell, these studies were performed with an accuracy and repeatability not previously obtainable or reported.

Fig. 4 is a block diagram of the experimental apparatus. A Uniphase 1105P 8 mW HeNe laser emitting at 632.8 nm supplied the incident light. The plane of polarization was controlled by a rotating zero order 1/2 wave plate. A rotating polaroid sheet removed any stray polarization components and was used to check the plane of polarization. We estimate that the incident linear polarization was at least 500:1 before diffraction by the SAW device. The SAW device was mounted on precision motorized positioning stages under control of a Newport 855C Controller. The linear y and z stages were accurate to 0.1 micron and the rotary stage in the xz plane had angular resolution to 0.001 degrees. Diffracted light was detected using an EG&G Optical Multichannel Analyzer (OMA) equipped with a 1254 silicon intensified vidicon detector. The detector has a 512 horizontal by 256 vertical element array in a 12.5 mm by 12.5 mm area. Each detector element is approximately 24 microns wide by 48 microns high. A Zenith Z-248 AT compatible computer controlled the entire experiment using IEEE 488 interfaces. Original software was written for automated data collection and storage.

The RF input to the SAW device came from a Hewlett-Packard 8657A signal generator (0.1 to 1040 MHz) amplified by an ENI 603L 40 dB RF power amp. The input signal strength was monitored by either a Hewlett-Packard 437B RF power meter or a Tektronix 11402 Digital Oscilloscope. The acoustic signal was applied continuously during data acquisition, although facilities for pulsed RF operation are available for future experiments. This apparatus consistently provides micron positioning repeatability and fully automated data acquisition. First, the SAW device was aligned manually to obtain maximum diffraction for a given acoustic power and y polarized light. Once the device was properly aligned for Bragg diffraction, the detector was scanned to find the point of peak diffracted beam intensity. If the detector signal was acceptable, the rotary stage was then reversed to -4.000 degrees and data acquisition began. A background light reading was acquired just before the RF was turned on for each data point. The background was then subtracted electronically from the data reading before storage in the OMA. The SAW device was rotated in 0.050 degree increments and the peak diffracted light intensity vs. incident angle was plotted. The waveplate and polaroid were then rotated together to the orthogonal polarization and the data acquisition started again. The vidicon had the unique advantage of being

able to track any movement of the diffracted beam vertically or horizontally during data acquisition. This improved the accuracy and repeatability of the measurement. The Photonics Research Cell has only been in operation for 1 year; these experiments represent the first original experimental research data published from this facility.

Experimental results over a broad range of acoustic frequencies are shown in Fig. 5. Since the intensified vidicon photon counting is sensitive to 2 photons per count, this is the most accurate data yet obtained for anisotropic SAW diffraction effects. Clearly, the diffraction efficiency for the two polarizations are not equal. The angular bandwidth broadening and midband intensity dip for z-polarized light are both evident. These features were observed over a bandwidth of 90-140 MHz.

Close inspection of Fig. 5 reveals that the relative magnitude of the intensity dip for z-polarized light increases slightly with increasing frequency. This is believed to be a scaling effect arising from the anisotropic Bragg angle equations presented earlier. Recall also that the theory predicts the relative magnitude of the dip should vary with acoustic signal strength. This was confirmed for the first time by studying diffracted light intensity vs. angle for a fixed acoustic frequency and varying RF signal amplitude into a 50 ohm termination. The results are summarized in Fig. 6. As expected, the intensity dip becomes smaller as the applied RF power decreases.

Conclusions

We have developed a theory to describe anisotropic SAW interaction, including the effects of incident light polarization, in collaboration with Rensselaer Polytechnic Institute. The SAW is modeled as a superposition of longitudinal and transverse bulk wave components, each of which may be analyzed separately by coupled mode theory. Computer simulations of this theory predict different diffraction efficiencies for y and z polarized light in lithium niobate, as well as a broader angular bandwidth and midband intensity dip for z-polarized light. These results have been confirmed by the Photonics Research Cell at West Point using a highly accurate computer controlled data acquisition system. These are the first repeatable measurements of this phenomena, as well as the most accurate. Good quantitative agreement was obtained between theory and experiment. Future research efforts include studies of optical wavelength dependence and diffraction in other material, particularly GaAs, for integrated optics applications.

Acknowledgement

of Mr. Charles Garvin, Mr. Joe Mait, and Mr. Fred Semendi of Harry Diamond Laboratories, Adelphi, MD.

References

1. P. Das, <u>Optical Signal Processing</u>, Springer-Verlag, to be published.
2. N. J. Berg and J. N. Lee, Editors, <u>Acousto-Optic Signal Processing</u>, Marcel Dekker, 1983.
3. P. Das, L. Milstein, "Surface acoustic wave devices", IEEE Comm. vol. 17, no. 5, p. 24-33, September 1979.
4. A. V. Scholtz, P. Das, D. M. Litynski, A. J. Urillo, "Coupled mode theory for the acousto-electro-optic effect", presented at the 1986 Optical Society of America Annual Meeting, October 1986 and to be published.
5. P. Zory and C. Powell, "Light diffraction efficiency of acoustic surface waves", App. Opt. vol. 10, no. 9, p. 2104-2106 (1971).
6. R. A. Davidson, C. DeCusatis, P. Das, "Optimization of acousto-electro-optic signal processing in anisotropic media", Proc. IEEE Ultrasonics Symposium, p. 495-500, October 1987.
7. D. G. Feitelson, <u>Optical Computing</u>, MIT Press, Cambridge, MA (1988).
8. S. W. Li, E. Bourkoff, N. J. Berg, C. Garvin, "Polarization effects in acousto-optic signal processors: theory and experiment", SPIE Proc. Analog and Optical Computing, p. 414-419 (1985).
9. D. K. Schockley, C. Garvin, "Observed polarization dependence of the SAW acousto-optic interaction in $LiNbO_3$", SPIE Proc. 789, p. 145-148 (1987).
10. P. Das, D. M. Litynski, C. DeCusatis, "A theory of acousto-electro-optic interaction in anisotropic media, including polarization effects", presented at the 1987 Optical Society of America Annual Meeting, October 1987.
11. R. J. Pressley, Editor, <u>Handbook of Lasers with Selected Data of Optical Technology</u>, CRC Press, Cleveland (1971).
12. V. Ramachandran, "Electric field intensity distribution inside a $LiNbO_3$ crystal with different interdigital electrode configurations", J. Phys. D: Appl. Phys. 11, p. 1445-1449 (1978).
13. R. N. Spaight and G. G. Koerber, "Piezoelectric surface waves on $LiNbO_3$,", IEEE Trans. on Sonics and Ultrasonics, vol. SU-19, no. 4, October 1971, p. 237-238.
14. R. W. Dixon, "Acoustic diffraction of light in anisotropic media", IEEE Journal of Quantum Electronics vol. QE-3, no. 2, p. 85-93 (1967).
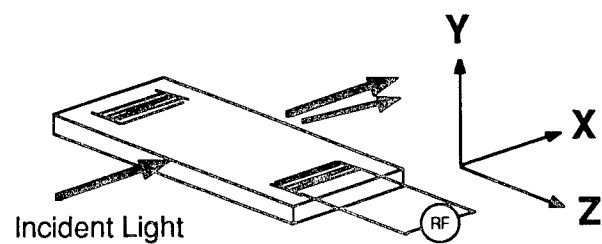
Figure 1   Crystal geometry and coordinate system
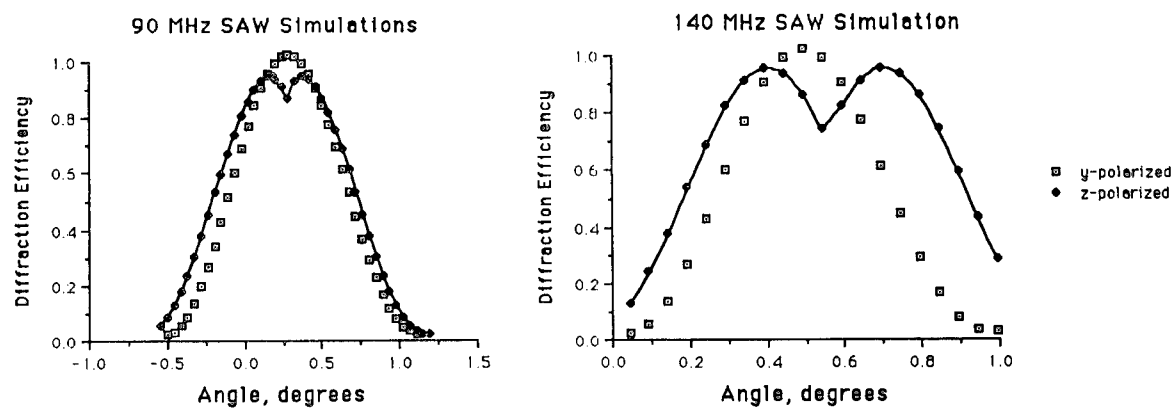
Figure 2   Computer simulation of SAW diffraction at 90 MHz and 140 MHz

7300 microns

400 microns

8000 microns

Mask design of split finger transducer on 180 MHz device (not to scale)

Each finger 6 microns wide, 3 microns between fingers, 26 fingers total
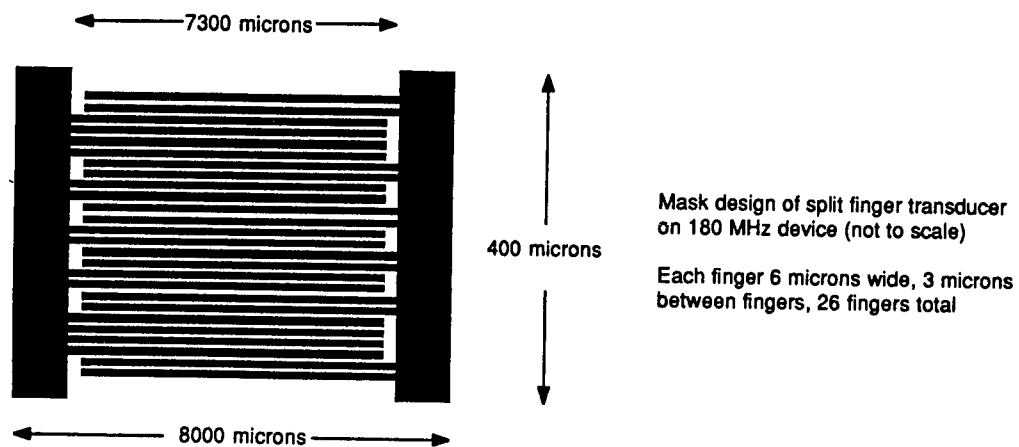
Figure 3   Split finger transducer design

1 / 2
Waveplate

Collimator

CYL
Lens

SAW Device
on Z, Y, and
Rotary Stage

CYL
Lens

Stop

HeNe

ND
Filter

Polaroid

Stop

Vidicon
Detector

RF
Amp

Vidicon
Controller

Motion
Controller

Signal
Generator

OMA

IEEE 488 Bus
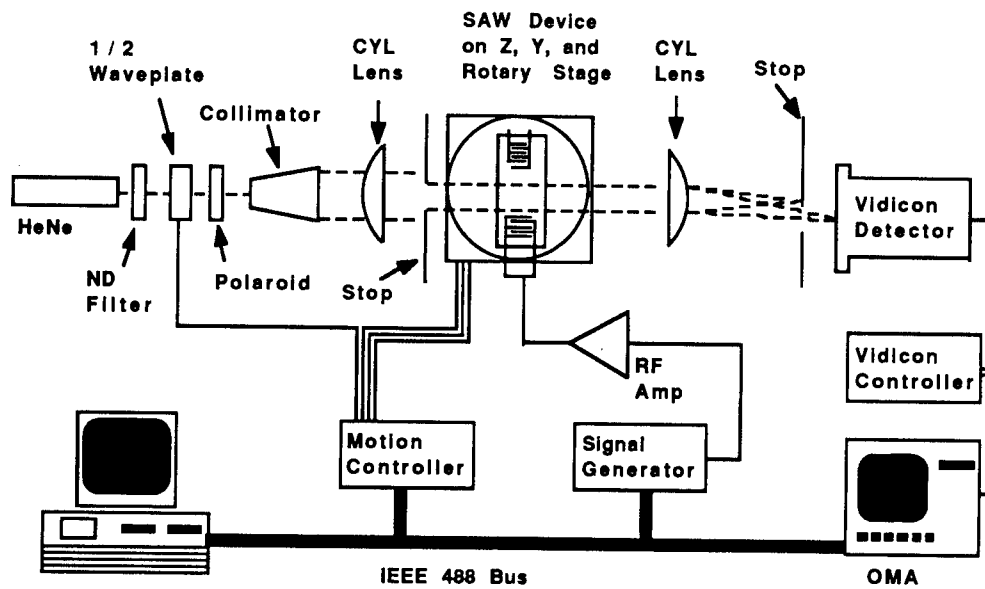
Figure 4   Block diagram of experimental apparatus
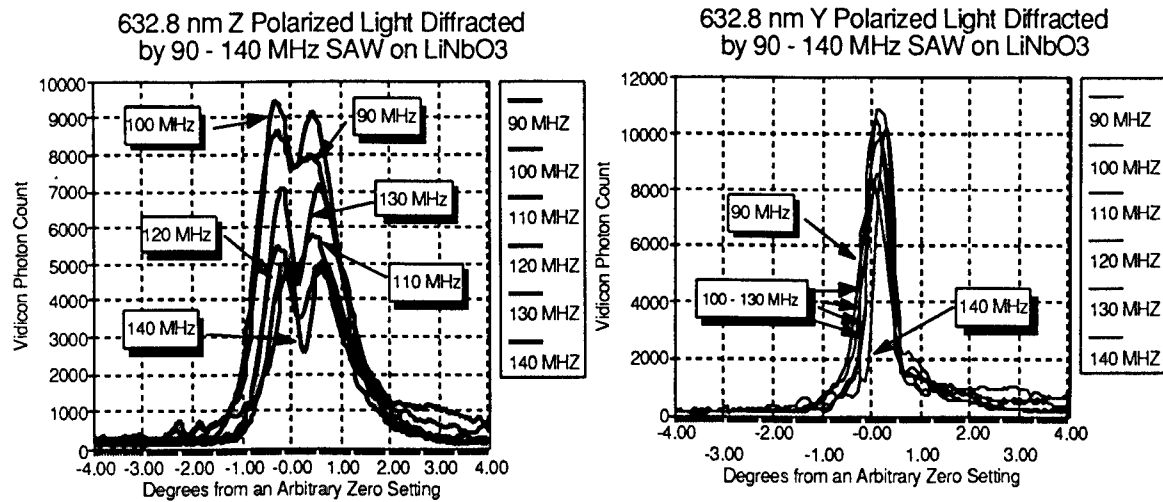
Figure 5   Experimental SAW diffraction at different acoustic frequencies
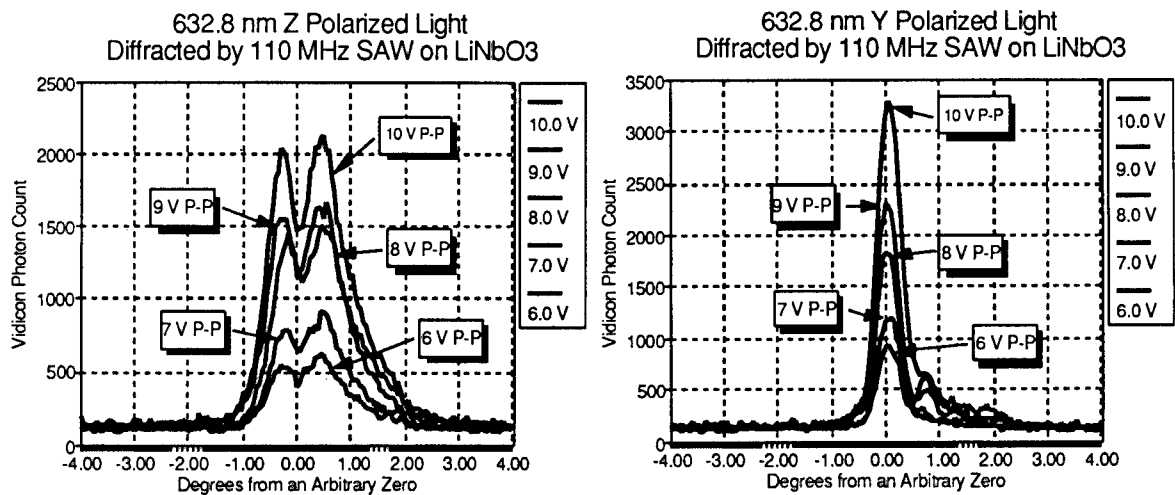
Figure 6   Experimental SAW diffraction for different acoustic voltages

All-Optical Recombination Time Studies in Wide Gap Semiconductors

John E. La Sala, LTC, EN
*Leo S. Hergenroeder, MAJ, FA
Richard L. Hansen, 2LT, EN

Department of Physics
United States Military Academy
West Point, New York 10996

INTRODUCTION. Semiconductors possess large third order non-linearities that have been exploited in optical limiting investigations (see Ref.1 and references contained therein). Group IV and III-V semiconductors have been most extensively studied as associated fabrication technologies are well developed and high quality samples are readily available. Precision doping and substitutional processes in these semiconductors are available to meet specific limiting and switching needs in the near infrared. The situation in the visible is not so well developed. The group IV and III-V semiconductors possess energy gaps that are too small to transmit the full visible spectrum. However, group II-VI semiconductors generally have wider gaps, and many are fully transmitting in the visible. In particular, ZnSe and ZnO have band gaps of 2.7 eV and 3.2 eV at room temperature[2], with cut-off wavelengths of 460 nm and 390 nm, respectively. High quality ZnSe is readily available in polycrystalline form prepared by the CVD process.[3] ZnO is grown as a single crystal.[4]

ZnSe has recently been used in the demonstration of a passive, picosecond optical power limiter.[5] The device used two-photon absorption leading to Drude type self-defocusing of an intense laser pulse focused inside the crystal. Effective limiting was achieved with 30 ps pulses, but for longer pulses (ns) competing thermal self-focusing is expected to dominate, destroying the crystal. A major factor in determining the pulse length that can be safely limited is the free carrier recombination time. Generally the longer the recombination time the longer defocusing effects successfully compete with thermal effects.

Free carrier recombination rates for bulk ZnSe and ZnO are not well known. No reported rates for recombination in ZnO are known to the authors. A recombination time on the order of 1 ns for CVD ZnSe is typically assumed.[1] Based on this figure recombination is usually ignored when modelling short (30 ps) pulses propagating through the crystal.[1] However, carrier-dependent recombination may result in much faster recombination at the very high carrier concentrations reached in optical limiting applications. It is the purpose of the present research to measure recombination rates at both high and low carrier densities in CVD ZnSe and single crystal ZnO.

LASALA, HERGENROEDER & HANSEN
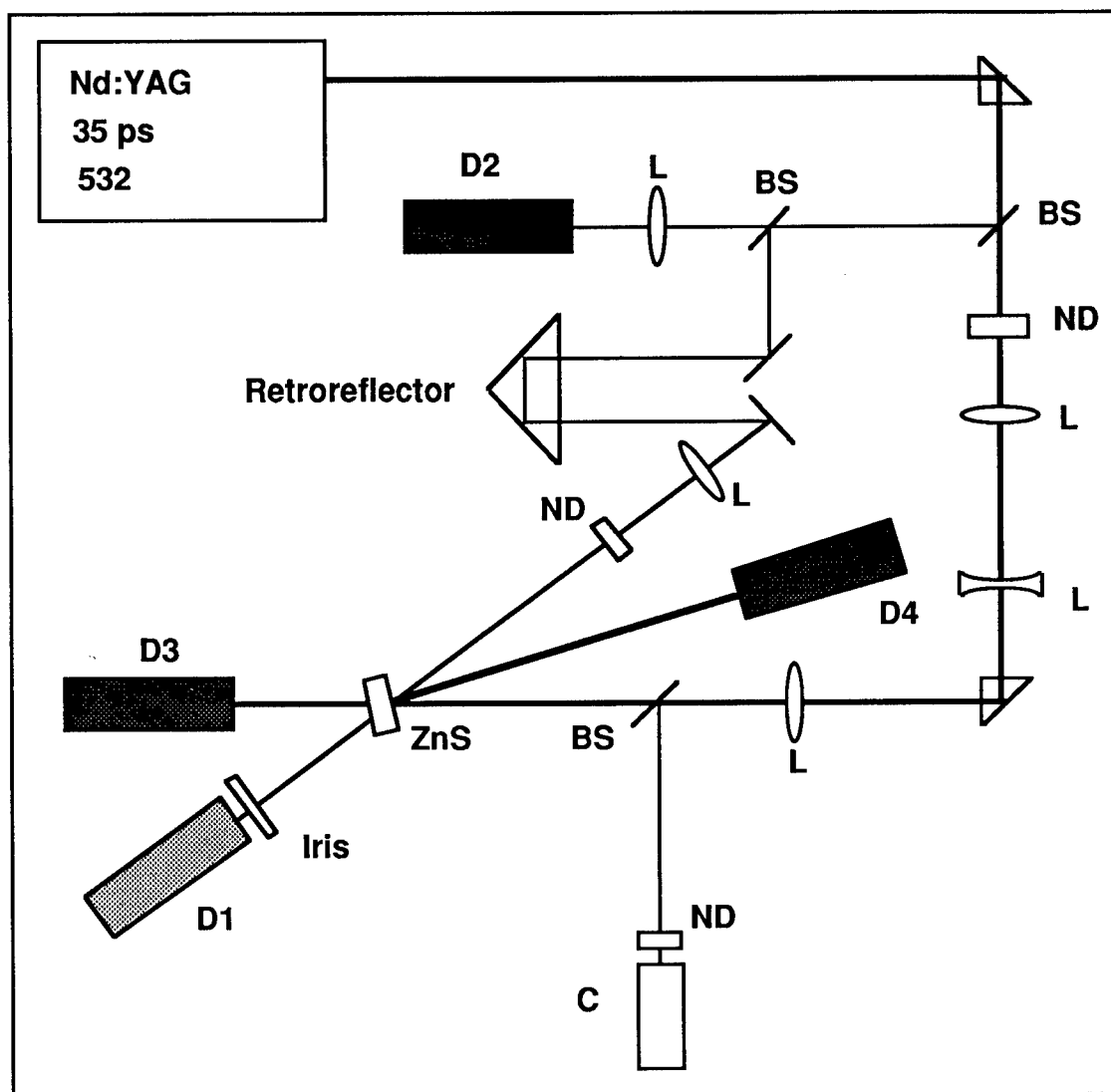
Figure 1. Layout of pump-probe experiment.

EXPERIMENT. Short (35 ps FWHM) pulses from a 10 Hz mode-locked Nd:YAG laser which have been frequency doubled to 532 nm are switched at a low repetition rate into a pump-probe beamline, shown in Figure 1. The beam is split into pump, probe and sampling pulses, all of which are vertically polarized for the initial experiments. The pump beam is brought to approximately a 1 mm diameter ($1/e^2$) focus at the front face of a 1 cm^2 ZnSe sample, which is tilted slightly to direct the reflected beam into a Laser Precision pyroelectric

LASALA, HERGENROEDER & HANSEN

energy detector D4. The soft focus insures that the pump remains collimated in the crystal, which is 3.0 mm thick. The reflection from the back face of the ZnSe crystal is also captured by reflection detector D4. The transmitted pump pulse enters an identical pyroelectric detector D3.

A beamsplitter located upstream of the crystal directs a small portion of the pump pulse into a CCD camera C1, located at the focus. This camera is one of four available in the Big Sky Multicam system, which is capable of capturing high resolution (25 μm) transverse beam profiles on a pulse by pulse basis.

The probe beam passes through a variable delay line, after which it also is focused at the crystal, with a spot size approximately three times smaller than the pump, but large enough to also remain collimated during its passage through the crystal. The probe crosses the pump at the front surface of the sample at an angle of 16°, which results in a crossing angle of 3.8° inside the high index (n = 2.7) ZnSe crystal. The reflected probe beam is not measured. The transmitted probe beam is captured by a Laser Precision silicon energy detector D1 capable of measuring energies from 2μJ to 20pJ.

The sampling detector D2 is calibrated to detectors D1 and D3 by removing the ZnSe sample. Detectors D2, D3 and D4 are mutually calibrated off line, and are found to agree within 5% of each other, which is the manufacturer's specification for pyroelectric and silicon detectors alike. The spectral response of the pyroelectric detectors is flat across the visible, while that of the silicon detector peaks in the near infrared, with a 33% residual response at 532 nm.

With the crystal in place, incident energies can be determined by scaling from the signal at D2. The typical incident pump energy used is 325 μJ and the incident probe energy is 120 nJ. For our 35 ps pulses these correspond to peak powers of 9.1 MW and 3.4 kW, respectively. The pump is focused to a near gaussian with a spot size $\omega = 0.40$ mm as measured by camera C1 during the experiment. This results in a peak intensity of 3.7 GW/cm^2 and a peak fluence of 0.13 J/cm^2, safely below the damage threshold of ZnSe which is approximately 1 J/cm^2. The high peak intensity of the pump is responsible for all nonlinearities observed. The probe is focused to a near gaussian of spot size one third that of the probe, resulting in a peak intensity of 0.012 GW/cm^2, well below onset of self-induced nonlinearities. The probe transverse profile is measured by replacing the ZnSe crystal with another camera C2. This camera is also used to align pump with probe, and to align probe transport optics to insure that it does not walk across the face of the crystal as the delay is changed.

The probe delay line employed a Melles Griot retroreflector and a Daedel translation stage, which provides a total delay from -150 ps (probe arriving before pump) to +1000 ps. Resolution of the translation stage is 10 μm, or 0.066 ps delay.

LASALA, HERGENROEDER & HANSEN

MODEL. To study recombination rates, an initial concentration of excess free carriers is produced by two photon absorption (TPA) of the pump according to the relation[6]

$$\frac{\partial N}{\partial t} = \frac{\beta_2 I^2}{2\hbar\omega} - rN + \text{diffusion term}$$ (1)

where N is the excess carrier density, β_2 is the TPA coefficient (5.5cm/GW for ZnSe[2]), proportional to the imaginary part of the material $\chi^{(3)}$, I is the pump intensity and r is the excess carrier recombination rate, which is in general a function of N. The diffusion term is not important for the spot sizes used, since recombination will occur on a time scale orders of magnitude shorter than the diffusion time (a few ms). In calculating the excess carrier concentration produced by the pump we ignore recombination during the passage of the pump. The validity of this assumption is examined in the results section.

The pump pulse will be depleted as is passes through the crystal according to[7,8]

$$\frac{dI}{dz} = -\alpha I - \beta_2 I^2 - \sigma_{eh} NI$$ (2)

where α is the linear absorption coefficient, measured by us to be 0.62 cm^{-1} (compared to 0.5 cm^{-1} reported in Ref. 8), β_2 is the TPA coefficient mentioned above and σ_{eh} is the combined electron and hole free carrier absorption (FCA) cross section. We calculate $\sigma_{eh} \cong 2 \times 10^{-18}$ cm^2 from the Drude model (see for example Ch. 1 Ref. 9) using the mobility and effective mass data for ZnSe.[10] In principle we may simultaneously solve equations (1) and (2) if we take the pump to be a gaussian in both time and space, and assume a form for the recombination rate. To simplify this complex calculation we decouple the equations by ignoring recombination during the pulse in (1) and ignoring the FCA term in (2). The first assumption is equivalent to assuming $1/r \gg \tau_p$, where τ_p is the FWHM of the (gaussian) pulse in time, an assumption we test with the data. The second assumption is validated by computing the maximum excess carrier concentration generated by the pump. An estimate is obtained by solving (1) for N at the entrance to the crystal after the full pulse has passed (ignoring depletion) resulting in N $\cong 1.6 \times 10^{18}$ cm^3. This gives an effective attenuation of σ_{eh}N $\cong 3$ cm^{-1}, compared to β_2I $\cong 16$ cm^{-1}. Therefore, we expect FCA to play only a minor role in pump depletion through the crystal.

The solution of (2) under the above assumptions is easily found by direct integration:

$$I = \frac{I_o(1-R)e^{-\alpha z}}{1 + \beta_2 I_o(1-R)(1-e^{-\alpha z})/\alpha}$$ (3)

LASALA, HERGENROEDER & HANSEN

where I = I(z,r,t) is the intensity at a depth z in the crystal, a distance r from the propagation axis and at time t during the passage of the pulse. I_o is the incident intensity, which is a function of time and transverse coordinates, and R is the reflection loss at the front surface.

In the above analysis diffraction of the beam inside the crystal has been neglected, since we use a collimated beam. In fact, the optical Kerr effect, resulting from the real part of the $\chi^{(3)}$, will produce a self-focusing of the beam, and the free carriers will contribute to a Drude type self-defocusing of the beam. This latter effect is exploited in some optical limiting applications. A rough calculation indicates that neither effect will significantly affect the beam inside the crystal, although subsequent propagation may be affected due to accumulated phase curvature.

With the above solution for I(z) the excess free carrier concentration may be calculated at points within the crystal by integrating equation (1) over the duration of the pump pulse, neglecting recombination during the passage of the pump. We assume a form for I_o as follows:

$$I_o(r,t) = I_o e^{\frac{-2r^2}{\omega^2}} e^{\frac{-t^2}{2\sigma_t^2}}$$

and compute N(z) by carrying out the time integration along r = 0 only, since the probe is focused more tightly that the pump in order to sample only points near the axis. The resulting integral

$$N(z) = \frac{\beta_2 I_o^2 (1-R)^2 e^{-2\alpha z}}{2\hbar\omega} \int_{-\infty}^{+\infty} \left[1 + \frac{\beta_2 I_o (1-R)}{\alpha}\left(1 - e^{-\alpha z}\right)e^{-\frac{t^2}{2\sigma_t^2}} \right]^{-2} e^{-\frac{t^2}{2\sigma_t^2}} dt$$

must be solved numerically, assuming values for incident peak intensity of $I_o = 3.7$ GW/cm^2, intensity reflectivity R = 0.196 for 5° angle of incidence, and values of linear and two-photon absorption coefficients mentioned above. The result of this integral is shown in Figure 2.

The probe pulse, arriving at time Δt after the pump pulse has passed, will sample the free carrier concentration and suffer attenuation due to FCA, as well as linear absorption, as it passes through the crystal. For a fixed delay Δt sufficiently large that pump and probe do not overlap in time, the transmission of the probe is obtained by solving equation (2) without the nonlinear term:

$$I_{pr}(\Delta t) = I_{o,pr}(1-R)^2 e^{-\alpha d} \exp\left(-\sigma_{eh}\int_0^d N(z,\Delta t)dz\right) \qquad (4)$$

where N has been represented as a function of z and Δt to make recombination effects evident. If we assume recombination occurs on a time scale long compared to the probe length (35 ps), then probe fluence transmission will scale with the intensity transmission as in equation (4).

LASALA, HERGENROEDER & HANSEN

Further, if we assume the probe beam is sufficiently smaller than the pump in spot size and that walk-off of probe and pump due to the finite crossing angle in the crystal (3°) is small, then the energy transmission will also scale as the intensity. Under these assumptions, probe energy transmittance will provide direct sampling of the free carrier population.
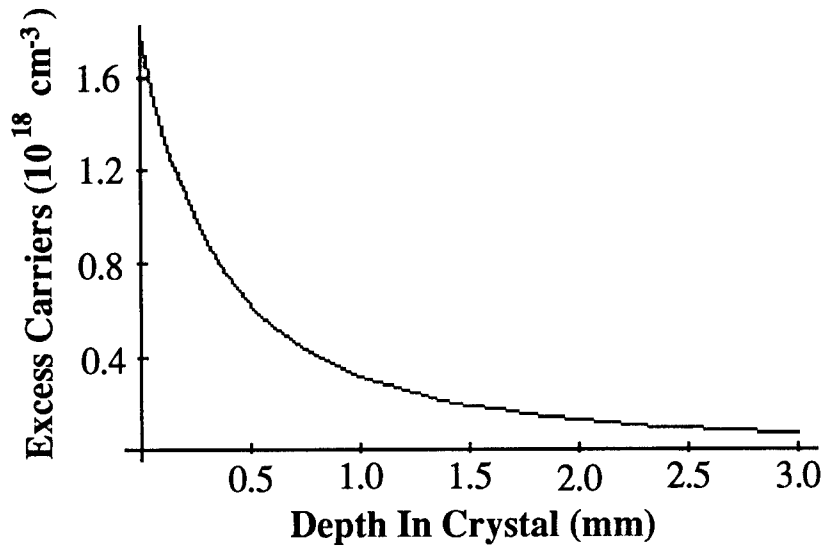
Figure 2. Profile of excess carrier density N(z) due to passage of pump pulse through the ZnSe sample. The sample is 3.0 mm in depth. Calculation assumes no free carrier recombination occurs during the passage of the pump pulse (35 ps).

The maximum probe attenuation due to FCA will occur at $\Delta t = 0$, corresponding to the free carrier population profile shown in Figure 2. The integrated FCA exponent for this profile is found to be 0.21, assuming the previously mentioned value for σ_{eh}. This is sufficiently small to motivate an expansion of the exponent in (4):

$$T(\Delta t) \equiv \frac{E_{pr}(\Delta t)}{E_{o,pr}} \cong (1-R)^2 e^{-\alpha d}\left(1 - \sigma_{eh}\int_0^d N(z,\Delta t)dz\right), \qquad (4a)$$

where $E_{pr}(\Delta t)$ represents the probe energy transmitted with the delay set at Δt, $E_{o,pr}$ is the energy incident at the first surface of the crystal and $T(\Delta t)$ defines the energy transmission at delay time Δt. Experimentally, we measure the probe energy transmission with the pump

LASALA, HERGENROEDER & HANSEN

alternately blocked (T_b) and unblocked (T_u). Dividing the T_u by T_b removes the linear effects (reflection and attenuation) leaving a quantity T_{ub} that is a measure only of FCA attenuation:

$$T_{ub}(\Delta t) \cong 1 - \sigma_{eh} \int_0^d N(z, \Delta t) dz,$$

or

$$A_{ub}(\Delta t) \equiv 1 - T_{ub}(\Delta t) \cong \sigma_{eh} \int_0^d N(z, \Delta t) dz. \tag{5}$$

The attenuation A_{ub} is the experimentally measured quantity.

The excess free carrier recombination rate may be modeled phenomenologically by the following:[11]

$$rN = \frac{1}{\tau_R} N + BN^2 + CN^3 \tag{6}$$

The linear term represents non-radiative and impurity-mediated recombination, the square term radiative recombination, and the cubic term Auger recombination.[11] If the recombination is dominated by the linear term, the probe absorption A_{ub} will follow a simple exponential law in delay time:

$$A_{ub}(\Delta t) \cong e^{-\frac{\Delta t}{\tau_R}} \sigma_{eh} \int_0^d N(z, 0) dz. \tag{7}$$

If the higher order terms are important, $N(z, \Delta t)$ is not separable in z and Δt, and a more complex analysis of data will be required.

RESULTS. Figure 3 represents the data obtained to date. The figure shows probe attenuation A_{ub} as defined in equation (5) measured against pump-probe delay Δt. Zero delay is the nominal overlap of peak arrival times. Random error in delay measurement is negligible on this scale. Vertical error bars represent scatter in the data obtained at each delay position. Pulse to pulse energy variations from the laser result in variations in both pump intensity (hence carrier concentration) and probe intensity. A window from 280 - 350 µJ incident pump energy is selected to give the highest number of data points with the minimum spread in pump intensities. This range of pump energies gives a range of excess free carriers of approximately ±20%. We verify that the probe intensities remain small enough in this range to behave linearly in the absence of the pump.

LASALA, HERGENROEDER & HANSEN
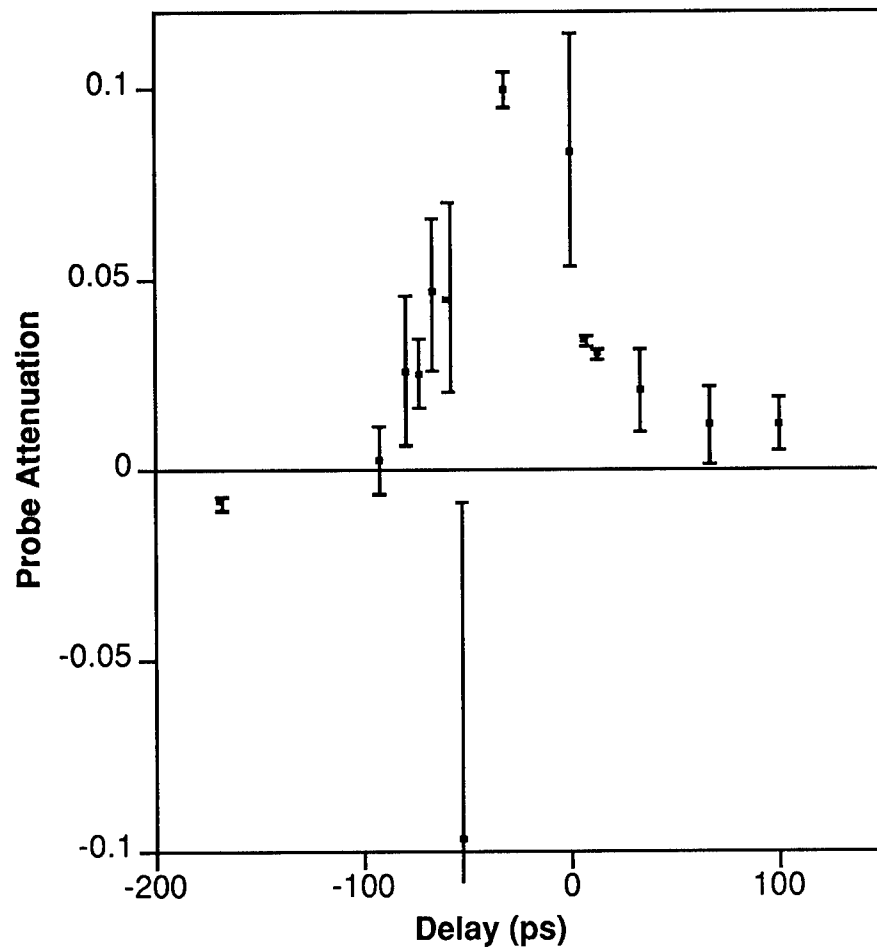
Figure 3. Normalized probe attenuation A_{ub} versus pump-probe delay. Positive delay represents the probe arriving later than the pump. Pump and probe lengths are 35 ps FWHM.

LASALA, HERGENROEDER & HANSEN

The first observation is that the magnitude of the attenuation A_{ub} is about a factor of two smaller than expected (0.19). Another obvious feature is that the peak is occurring somewhat earlier than the nominal zero delay point, in the vicinity of -30 ps. However, the location of zero delay is based on physical measurement of path lengths, which carries a systematic error as large as a 1 cm, or 33 ps. The asymmetric shape of the curve suggests FCA is occurring as expected, since it is an integrated effect. However, as a precaution against other attenuation processes that might be occurring while both pump and probe are overlapping in time, we initially concentrate on the latter five points as representative of probe attenuation due only to FCA, well after it has separated temporally from the pump.
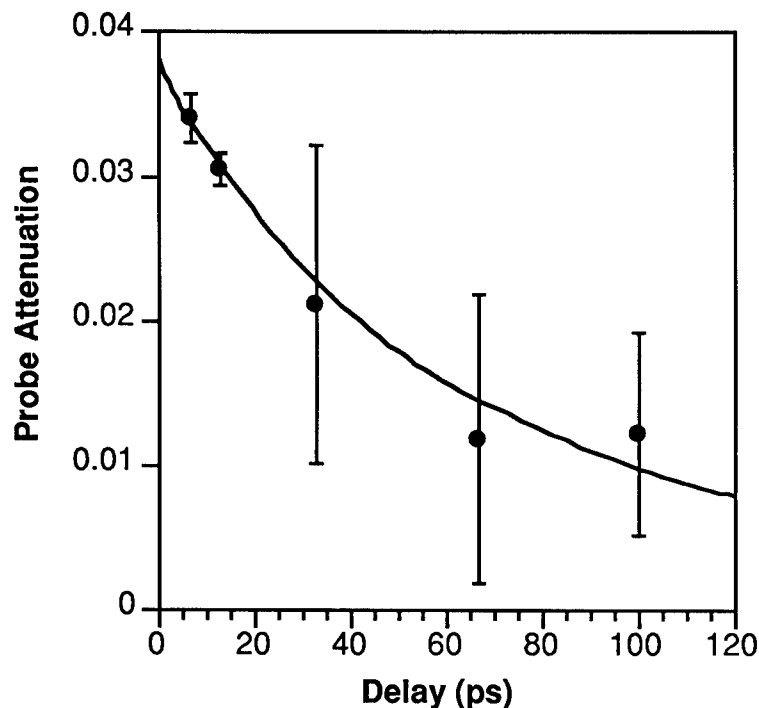
Figure 4. Region of Figure 3 that represents excess carrier recombination as a function of time between pump and probe. The solid curve represents the best fit of the data to the equation $\dfrac{\partial N}{\partial t} = -\dfrac{N}{\tau_R} - BN^2$.

LASALA, HERGENROEDER & HANSEN

In figure 4. we show a three parameter fit to the recombination data, modelled with the first two terms of equation (6). The fit parameters are τ_R, B and N_0. We find that including the second term improves the fit over that achieved with the linear term alone. The results obtained are as follows:

$$\tau_R = (110 \pm 30) \text{ ps,}$$

$$B = (2 \pm 1) \times 10^{-8} \text{ cm}^3\text{/s,}$$

$$N_0 = (3 \pm 1.5) \times 10^{17} \text{ cm}^{-3}.$$

In making these estimates, we have assumed the time dependence of probe attenuation reflects the time dependence of free carriers a distance $d_{eff} \sim 0.5$ mm into the crystal. This is necessary since the recombination rate itself depends on N(z) when the second order term is retained in the model. The estimates of B and N_0 are most directly affected by this approximation. Proper treatment of the integrated effect will be deferred until a larger set of data is obtained. Nonetheless, this approximation is likely good to within a factor of 2 for estimating B and N_0. Due to this assumption and the sizable error bars in absorption values and lack of points at large delay, these estimates should be taken as only very preliminary ones.

Projection of the above fit to earlier times offers an approximate method to assess the importance of the third order term in the recombination model and the validity of the assumption that no recombination occurs on the time scale of the pump length. Based on the initial estimate of maximum probe attenuation A_{ub} of approximately 0.19 the peak probe attenuation due to FCA would occur at a time of -60 ps in Fig. 3, earlier by 30 ps than the apparent peak around -30 ps. This observation suggests that the third order recombination term may be important at carrier concentrations reached early in the pulse, and that significant recombination (a factor of two) is occurring on the time scale of the pump length itself. However, other possibilities, including imperfect overlap of the pump and probe within the crystal and the finite spot size of the probe, would also lower the maximum attenuation expected. If these two effects lower maximum attenuation by a factor of two, we project a peak close to that observed without invoking the third term, restoring confidence in the earlier assumption of no recombination during the pump pulse.

Interpretation of the data appearing before the recombination tail is clouded by the possibility that while pump and probe overlap temporally, processes in addition to FCA may occur, including TPA (one pump and one probe photon), transient energy transfer (TET) and bleaching of linear absorption.[12] In particular, the drop in attenuation below zero near -50 ps suggests either diminishing of linear attenuation and/or true gain in the probe. Initial estimates of probe transmission indicate more energy is reaching the probe detector at that delay than can be accounted for by reflection losses. The iris placed before the probe detector is effective in blocking scattered pump light, eliminating that possibility for apparent gain.

LASALA, HERGENROEDER & HANSEN

The TET process, a gain mechanism suggested in Ref. 12 that may be active in our polycrystalline sample, is a second order photorefractive effect only possible with short pulses. This process would be accompanied by and compete with TPA, which attenuates the probe. Interestingly, the strong dip observed in our data occurs earlier than the peak in absorption, an effect observed in GaAs by Smirl, *et al* in Ref. 12.

CONCLUSIONS. We have made initial measurements of free carrier recombination in CVD ZnSe and find that carrier dependent recombination occurs on a time scale significantly shorter than previously believed. We also observe evidence of degenerate two-beam coupling in ZnSe. Both of these results are presently being explored in greater detail. In particular, we are repeating the pump-probe experiment with probe and pump orthogonally polarized to minimize beam coupling for the purposes of isolating FCA effects, and extending those measurements to longer positive delay values. In addition, improvements in data acquisition automation will increase the precision of attenuation measurements. With the FCA contribution identified we will investigate the direct beam coupling effects observed while pump and probe overlap temporally.

We plan also to repeat the above measurements in additional samples of ZnSe to identify any sample-specific phenomena. When samples arrive we will perform similar measurements in ZnO, in which first order photorefractive effects may also appear.

LASALA, HERGENROEDER & HANSEN

REFERENCES:

1. D.J. Hagan, E.W. Van Stryland, Y.Y. Wu, T.H. Wei, M. Sheik-Bahae, A. Said, K. Mansour, J. Young and M.J. Soileau, "Passive Broadband High Dynamic Range Semiconductor Limiters, *Materials for Optical Switches, Isolators and Limiters,* M. Soileau, Ed., Proc. SPIE 1105, 103 (1989).

2. E. W. Van Stryland, M. A. Woodall, H.Vanherzeele and M.J. Soileau, Opt. Lett. **10**, 490 (1985).

3. CVD Inc. Woburn, MA.

4. Atomergic Chemetals Corporation, Farmingdale, NY.

5. D.J. Hagan, E.W. Van Stryland, M.J. Soileau, Y.Y. Wu, S.Guha, Opt. Lett. **13**, 315 (1988).

6. J.H. Bechtel and W. L. Smith, Phys. Rev. B **13**, 3515 (1976).

7. C.R. Pidgeon, B.S. Wherrett, A.M. Johnston, J. Dempsey and A. Miller, Phys. Rev. Lett. **42**, 1785 (1979).

8. E. W. Van Stryland, H.Vanherzeele, M. A. Woodall, M.J. Soileau, A.L. Smirl, S. Guha, T.F. Boggess, Optical Engineering **24**, 613 (1985).

9. N.W. Ashcroft and N.D. Mermin, *Solid State Physics,* Holt, Rinehart & Winston, NY (1976).

10. M.A. Omar, *Elementary Solid State Physics: Principles and Applications,* Addison-Wesley, Reading, Massachusetts (1975).

11. J.P. Zheng and H.S. Kwok, Appl. Phys. Lett. **54**, 1 (1989).

12. A.L. Smirl, G.C. Valley, K.M. Bohnert and T.F. Boggess, IEEE-JQE **24**, 289 (1988).

Novel Quantum-Well Geometries for
Optoelectronic Device Applications (U)

*R. P. Leavitt, Dr.
U. S. Army LABCOM
Harry Diamond Laboratories
Adelphi, MD 20783–1197
and
J. W. Little, Dr.
Martin Marietta Laboratories
Baltimore, MD 21227-3898

1. Introduction

The Army has a need for fast optoelectronic components and circuits for future optical signal processing systems. The development of such components and circuits would exploit the massive parallelism that is achievable with optical processing and optical computing. However, conventional optoelectronic components are not sufficient to meet the stringent requirements for such systems; in particular, the discrete nature of these devices does not lend itself readily to chip-level integration, which is necessary to meet expected data-rate requirements.

To this end, there has been a great deal of interest in the development of optoelectronic devices based on thin, multilayer III-V semiconductor technology. Advanced growth techniques such as molecular-beam epitaxy (MBE) and organo-metallic chemical vapor epitaxy (OMVPE) have permitted the growth of extremely thin films (with individual layer thicknesses approaching a single atomic layer) of semiconductors whose electronic and optical properties are radically different from those of bulk semiconductors. Devices made using such structures have been shown to exhibit superior characteristics in comparison with those of corresponding devices fabricated from bulk semiconductors. In addition, several different types of III-V optoelectronic devices (such as lasers, amplitude and phase modulators, waveguides, switches, and detectors) can be integrated at the chip level on a common substrate, such as GaAs.

(a)   (b)
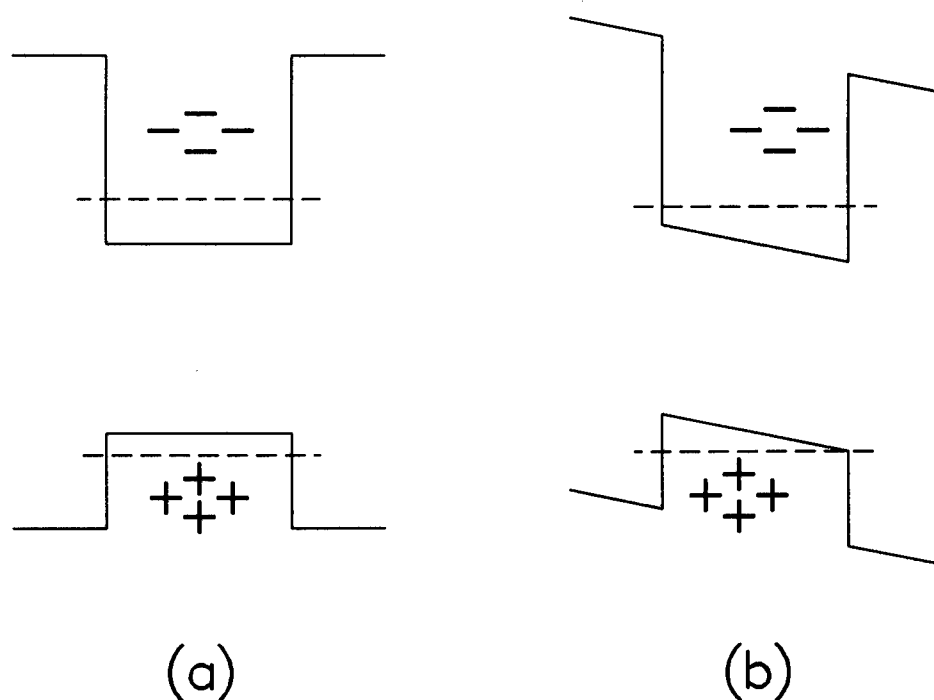
Figure 1. Band diagram for a rectangular quantum well. In (a), the electric field is zero, and there is no net polarization of an electron-hole pair in the well. In (b), an electric field has been applied, pointing to the left. The effect on the electron-hole pair is to induce a net polarization, with the mean positions of the electron and the hole being separated.

Quantum wells (consisting of a thin layer of semiconductor surrounded by layers of another semiconductor with a larger energy gap) are of particular importance to the development of optoelectronic devices. Structures containing quantum wells have been shown to exhibit electro-optic effects that are several orders of magnitude larger than those in bulk semiconductors; such effects have already resulted in the development of practical devices such as modulators, phase shifters, optically bistable logic elements, and current-injected diode lasers.

In Fig. 1 we show diagrams of the valence- and conduction-band edges (not to scale) for a quantum well formed from semiconductor system such as $GaAs/Al_xGa_{1-x}As$. In this system, both electrons and holes are confined by the potential well which results from the energy-gap difference of the two semiconductors. Because of quantum-mechanical effects, both electrons and holes exhibit non-zero confinement energies, which depend on the well width and also on the effective masses and

the energy-gap offsets (i.e., the depths of the valence- and conduction-band potential wells). These confinement energies are shown as dashed lines in Fig. 1. With no electric field (a), the potential profile is symmetric, and there is no net polarization of electrons and holes. In its lowest-energy state, an electron-hole pair placed in the well forms a bound state, called an *exciton*. Because of the spatial confinement, the binding energy of the exciton in a quantum well is considerably larger than in a bulk semiconductor. Consequently, the exciton is stable at room temperature, and the optical absorption features attributable to the exciton are more prominent in the absorption spectrum. In the quantum-well system, the fundamental absorption edge occurs at the energy needed to create the exciton, which is equal to the energy gap plus the sum of the electron and hole confinement energies minus the exciton binding energy.

When an electric field is applied to the well perpendicular to the layers, the electron and hole wave functions become polarized, i.e., there is a net spatial separation of the electron-hole pair. The electron and hole confinement energies decrease as a consequence, and the exciton binding energy decreases. The net effect is to decrease the energy of the optical absorption edge, which is called the *quantum-confined Stark effect* (QCSE).[1]

In this paper, we show that more complex structures consisting of a number of quantum wells coupled by thin barriers can exhibit electro-optic effects that are larger than (and potentially more useful than) the QCSE. We report the results of theoretical and experimental studies[2-10] of these coupled quantum-well structures grown by MBE. We have observed direct optical evidence of phenomena associated with coupling in asymmetric coupled quantum wells (ACQWs) made from a variety of III-V semiconductors. We have designed, grown, and characterized waveguide structures containing ACQWs that take advantage of enhancements in their below-bandgap optical properties for phase and intensity modulators. We show that strong coupling between the quantum wells in a short-period superlattice whose miniband widths are comparable to the band offsets leads to prominent spectroscopic features associated with Wannier-Stark localization, large effective blue shifts, and the possibility of Bloch oscillations.

2. Background

The principal advantage for device applications of coupled quantum-well systems over conventional uncoupled-well systems based on the QCSE described above is that the former provide more degrees of freedom for designing optimized structures. In structures having uncoupled, equal-sized wells, the only degrees of freedom are the width of the well and the total number of wells. In many applications (e.g., optical

modulators), these are predetermined by the system requirements, and there are no degrees of freedom left with which to optimize the design of the modulator. For example, in fiber-optics applications, which utilize InGaAs/InAlAs structures lattice matched to InP, it is desirable to have modulators that operate at either 1.3 or 1.55 μm. The width of a quantum well for a QCSE modulator is then constrained to be on the order of 3 nm for 1.3 μm or 7 nm for 1.55 μm. For wells this thin, the QCSE is weak, and large reverse biases must be applied to the modulator in order to obtain sufficient phase shift and/or modulation depth. Furthermore, although many studies of phase and intensity modulation reported in the literature have used multi-mode waveguide samples containing many quantum wells, in reality the military applications of interest may require single-mode waveguides, which must contain only a few wells to limit the overall thickness of the core region.

In the past several years, there have been publications by others on the optical properties of symmetric or slightly asymmetric coupled quantum-well structures.[11-13] In this paper, we are mainly concerned with the optical properties of asymmetric coupled-quantum-well systems in which the degree of asymmetry is relatively large. The band diagram for the system is shown in Fig. 2.

In Fig. 2(a), the magnitude of the applied electric field is zero. Electron and hole wave functions are localized in one well or the other; the lowest-energy states for the electrons and holes are localized in the wide well. At a certain value of the electric field (pointing to the left in the figure), the electron levels come into resonance, where strong coupling between the levels results in delocalization of both electron states, with the wave functions having substantial amplitudes in each of the wells. This coupling results in an avoided level crossing and an associated minimum energy splitting between the states that depends on the width and the height of the barrier between the wells. Figure 2(b) shows the system biased beyond the point where the electron levels couple. The wave functions are again localized in separate wells, but the lowest-energy electron level is now in the narrow well. As we show in this paper, the resonance of electron levels in the ACQW system leads to a wide variety of new phenomena (e.g., level splittings, appearance of new optical transitions, changes in the real and imaginary parts of the refractive index) that are applicable in a number of new optoelectronic devices.

3. Theory

The design of quantum-well structures for the experiments described here relies to a large degree on an accurate theoretical desciption of the electronic structure and optical properties of excitons in quantum wells. To this end, we have formulated a
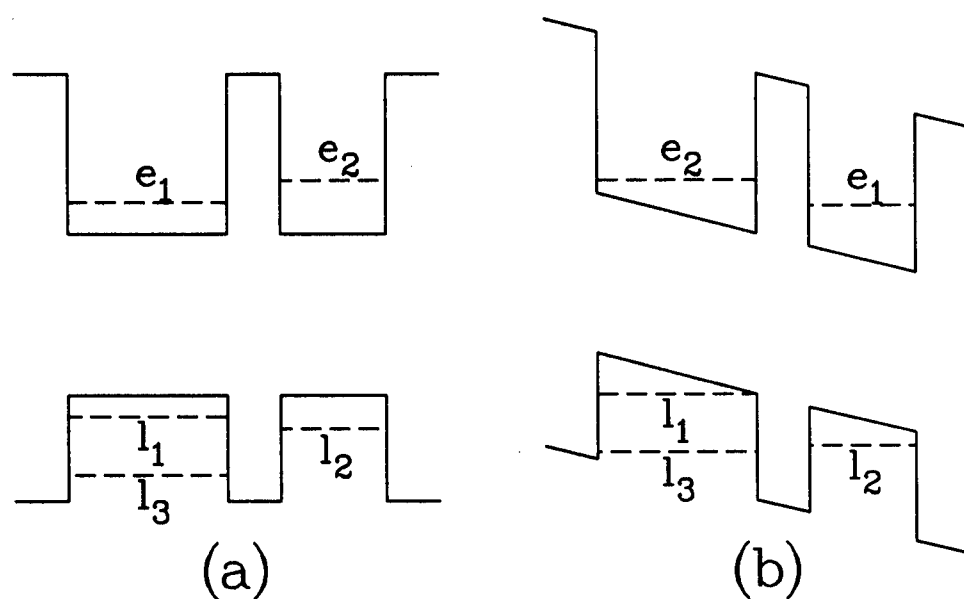
Figure 2. Energy-band diagram (not to scale) for the asymmetric coupled-quantum-well system for (a) zero field and (b) a field for which the electron levels are beyond resonance. The electron and light-hole levels are labeled $e_{1,2}$ and $l_{1,2,3}$, respectively, where the subscripts are in order of increasing energy. Not shown are the heavy-hole levels, referred to in the text as $h_{1,2}$.

theoretical model that accounts for several salient features of the quantum-well problem:

- Confinement energies of the electrons and holes are calculated by using a single-band effective-mass approximation. This approximation accounts for effects of electric fields by using the exact Airy-function solutions to the effective-mass Schrödinger equation with a linear potential, together with appropriate boundary conditions (continuity of the wave function and the probability current density) at each interface in the problem. A scattering phase-shift treatment is used to determine the eigenvalues of the quasi-bound states in an electric field.[14]

- The single-band treatment of confinement energies has been extended to include the effects of band nonparabolicity by extending a previously developed empirical two-band model[15] to quantum wells and superlattices in an electric field. This extension allows a more accurate treatment of the confinement energies in real systems.[16]

- Optical transition probabilities are determined by numerically calculating the overlap integrals between electron and hole wave functions.

- We have developed a new formalism[7] for calculating the binding energies of excitons in quantum wells and superlattices. In this formalism the binding energy is given as a simple integral of the squares of the electron and hole wave functions multiplied by a prescribed function. This development makes possible the rapid and accurate calculation of binding energies in complex quantum-well structures.

All the features described above have been incorporated into a single computer program that can be used to design quantum-well and superlattice structures for optoelectronic devices. From the results, one can determine in advance the optical spectrum and (through Kramers–Kronig relations) the index of refraction of a quantum-well structure as a function of light wavelength and electric field. This computer program was used to design the structures described below.

4. Photocurrent Spectroscopy of ACQW Structures

Devices based on electric-field effects in quantum wells and superlattices require fabrication of *p-i-n* diodes. This device geometry allows a reverse bias to be applied via contacts to *p*- and *n*-type regions of the diodes; the electric field is dropped entirely across the intrinsic (*i*) region containing the quantum wells, which is sufficiently thin (~ 1 μm) to produce high electric fields across the quantum wells at relatively low bias voltages. It is useful to characterize such structures with photocurrent (PC) spectroscopy, in which one measures the light-induced current in the diode as a function of the wavelength of the light for a number of different reverse biases. The basic experimental configuration is shown in Fig. 3. Chopped light from a tungsten filament lamp is dispersed by a monochromator and focused onto the sample, and a current-sensitive preamplifier and a lock-in amplifier are used to detect the photocurrent.

The observed PC in a p-i-n diode structure is given by[17]

$$I(\lambda) = \frac{e\lambda W(\lambda)}{hc}(1 - r)\left\{1 - \exp[-\alpha(\lambda)L]\right\}\eta(\lambda), \tag{1}$$

where $I(\lambda)$ is the collected photocurrent, $W(\lambda)$ is the power of the light incident on the sample, r is the reflectivity of the sample-air interface, $\alpha(\lambda)$ is the absorption coefficient of the intrinsic region of the diode, L is the intrinsic region thickness, and $\eta(\lambda)$ is the carrier collection efficiency of the diode. As long as $\eta(\lambda)$ is not a strong function of the light wavelength λ, the PC is proportional to the absorption spectrum of the diode's intrinsic region.
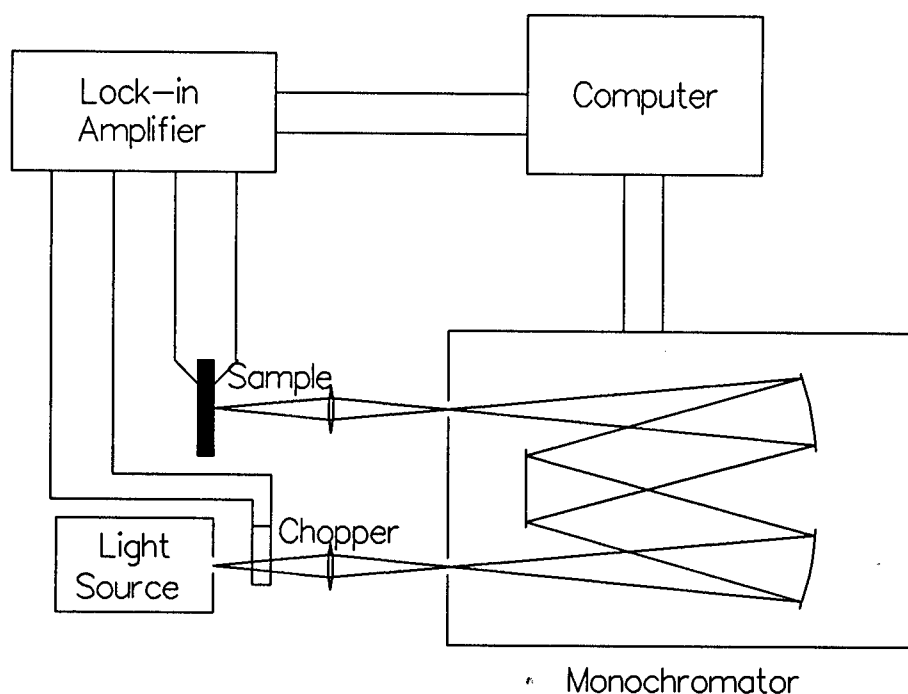
Figure 3. Schematic diagram of the experimental setup for photocurrent measurements.

Using the PC technique greatly simplifies sample processing. There is no need to remove the substrate (since the light does not need to be transmitted through the sample), and diodes are formed simply by depositing a blanket n-type (Au/Sn/Au) contact on the back surface of the wafer and etching mesas (on the order of 5 μm in height) with p-type (Au/Cr/Au) ring contacts. Typically, many diodes are processed on a single piece of material, and current-voltage characteristics are measured at room temperature to identify diodes with good electrical properties (high reverse-bias breakdown voltages, low leakage currents, etc.).

We have used room- and low-temperature PC measurements to characterize the optical transitions in a number of quantum-well and superlattice structures. As an example,[5] we consider here an ACQW structure containing 29 periods of 50- and 75-Å $In_{0.53}Ga_{0.47}As$ quantum wells separated by a 16-Å $In_{0.52}Al_{0.48}As$ barrier. The band diagram for the system is similar to that shown in Fig. 2. In Fig. 4, we show the low-temperature (8-K) photocurrent spectra for this sample for several values of the reverse bias. Because of the thin barrier, the coupling between electron states in the two wells is very strong, and small changes in bias result in significant changes in the spectra. The electron-state resonance occurs at about 5 V; beyond resonance, features
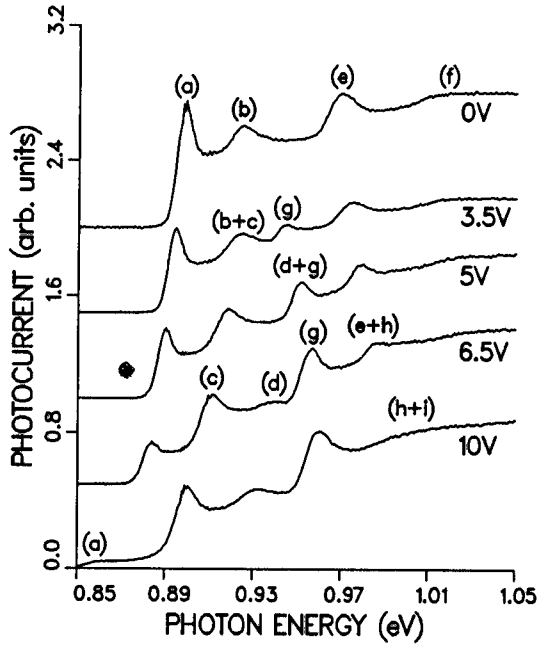
Figure 4. Low-temperature (8-K) photocurrent spectra of a sample containing 29 periods of 7.5- and 5.0-nm $In_{0.53}Ga_{0.47}As$ quantum wells coupled by a 1.6-nm $In_{0.52}Al_{0.48}As$ barrier, for several values of reverse bias. Transitions are labeled as follows: (a) h_1e_1; (b) l_1e_1; (c) h_1e_2; (d) l_1e_2; (e) h_2e_2; (f) l_2e_2; (g) h_2e_1; (h) l_2e_1; (i) l_3e_1. The spectra are offset vertically for clarity.
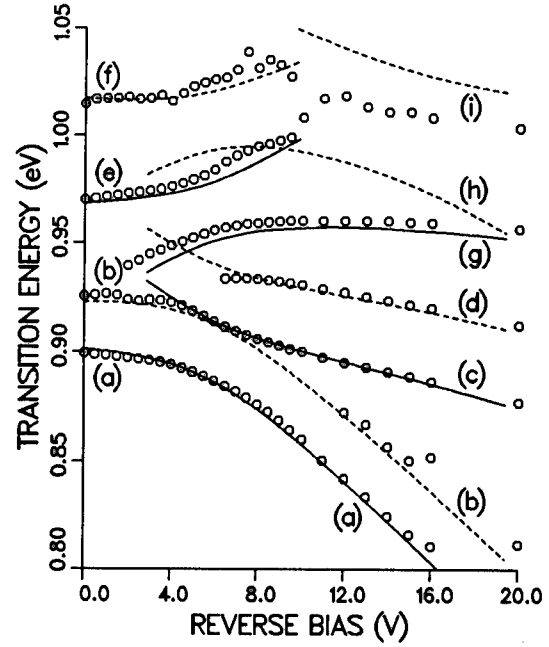
Figure 5. Comparison of calculated and observed (at 8 K) transition energies for the ACQW structure. Transitions are labeled as in Fig. 4. Solid lines indicate heavy-hole transitions, and dashed lines are the light-hole transitions.

of particular interest are the inter-well transitions (h_1e_1 and l_1e_1; see Fig. 2), which have substantial oscillator strengths that persist over a wide range of reverse bias (see, e.g., the shoulder labeled (a) in the bottom trace in Fig. 4). These transitions, as we shall see, are responsible for substantial enhancements in the below-bandgap optical properties of waveguide structures containing ACQWs.

Figure 5 shows the peak energies (symbols) of several excitonic transitions (such as those shown in Fig. 4) plotted as functions of reverse bias. Also shown are the calculated transition energies for heavy-hole (solid lines) and light-hole (dashed lines) transitions that are expected to have significant oscillator strengths. The calculations

were performed as described in Section 3. We predict an avoided crossing of the l_2 and l_3 hole levels [see Fig. 2(b)] at about 12–14 V, but the experimentally observed feature [labeled $(h+i)$ in Fig. 4] is too broad to be resolved. Otherwise, the agreement between theory and experiment is excellent, considering that the material parameters that enter into the calculations are not as well established for the $In_{0.53}Ga_{0.48}As$ system as for the $GaAs/Al_xGa_{1-x}As$ system.

Our PC spectroscopy work, including that described here, has shown that it is possible to obtain large shifts in the energies of excitonic absorptions in ACQW structures. The next section shows how these large electro-optic effects can be used in amplitude and phase modulators in a waveguide geometry.

5. Characterization of ACQW Waveguides

Because of the Kramers-Kronig relations, changes in the absorption spectra of quantum wells with an applied electric field also lead to changes in the refractive index. This physical principle can be applied to develop an electro-optic phase and/or amplitude modulator in which ACQW structures are contained in the core region of a single-mode semiconductor waveguide. We have fabricated and tested several waveguide modulators containing ACQW structures using both $GaAs/Al_xGa_{1-x}As$ and $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ quantum wells. Here we describe our results in the $GaAs/Al_xGa_{1-x}As$ material system.

The sample used for this study[4] consisted of seven periods of the ACQW structure embedded in a *p-i-n* diode waveguide. The layers and thicknesses comprising one period were: an 85-Å GaAs wide well, a 21-Å AlGaAs barrier, a 43-Å GaAs narrow well, and an 85-Å AlGaAs barrier between the pairs of coupled wells. With this ordering, an applied reverse bias could bring the electron subbands of the two wells into resonance, while the lowest energy light- and heavy-hole subbands of the wide well remained isolated from all other hole subbands. The cladding regions consisted of 1.5-μm $Al_{0.32}Ga_{0.68}As$ doped n^+ and p^+ (below and above, respectively) to 0.65 μm from the guiding (multiple-quantum-well) region. The result was a single-mode (slab) waveguide with a 0.164-μm-thick guiding region embedded within a 1.46-μm-thick intrinsic region.

The measurement system, shown in Fig. 6, consisted of a Mach-Zehnder interferometer with a pair of microscope objectives in one arm that end-fire-coupled the light from an Ar^+-laser-pumped Styryl-9M dye laser into the cleaved facet of a slab ACQW waveguide and collimated the emerging radiation. The interferometer output pattern was detected with a 0.3-mm-diameter Ge photodiode. The electrical bias was applied to the sample as a 500-Hz square-pulse train with a 50-percent duty cycle, and
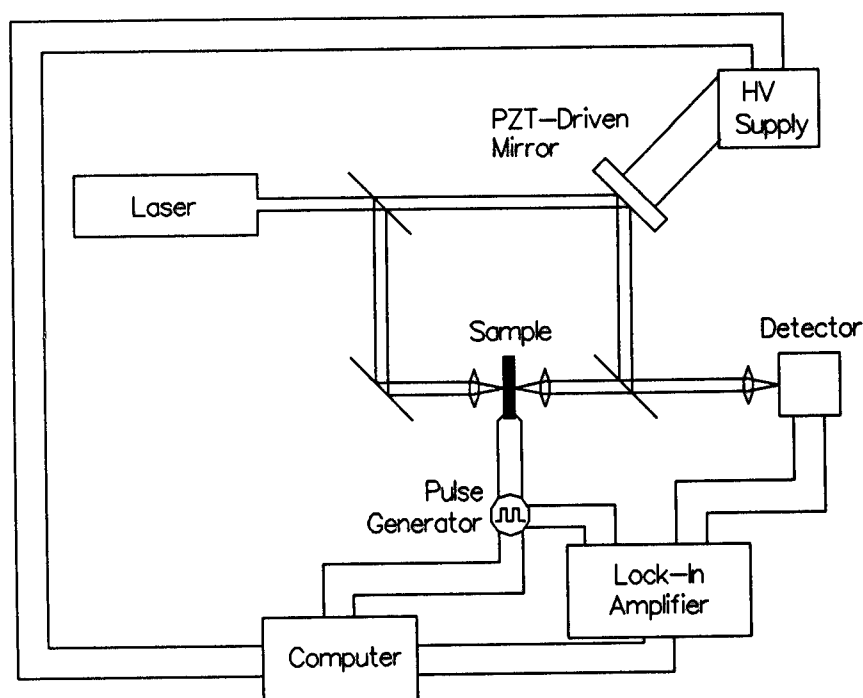
Figure 6. Schematic diagram of the Mach-Zehnder interferometer used for the waveguide experiments.

a boxcar averager was used to temporally separate the "bias-on" and "bias-off" signals. A piezoelectric transducer was used to slowly move the reference-arm mirror, tracing several periods of two interferograms (with and without bias applied to the sample).

In this work, the electric vector of the incident light was polarized perpendicular to the sample layer interfaces (i.e., TM polarized). With this polarization, only light-hole-to-electron transitions affect the optical properties of the guide, thus simplifying the interpretation of the observed phenomena.[18] In Fig. 7, we show the results of electro-refraction measurements made at four laser wavelengths, corresponding to energies of 45, 60, 84, and 107 meV, respectively, below the energy of the light-hole exciton, which occurs at about 835 nm at zero applied bias. A distinct feature is apparent (superimposed on a monotonically increasing background) in each of the curves in Fig. 7, which moves to higher biases and becomes weaker as the wavelength increases. The actual phase shifts observed in our 600-μm-long slab waveguide were as large as 300 degrees (at a wavelength of 861 nm and with an applied reverse bias of 15 V).
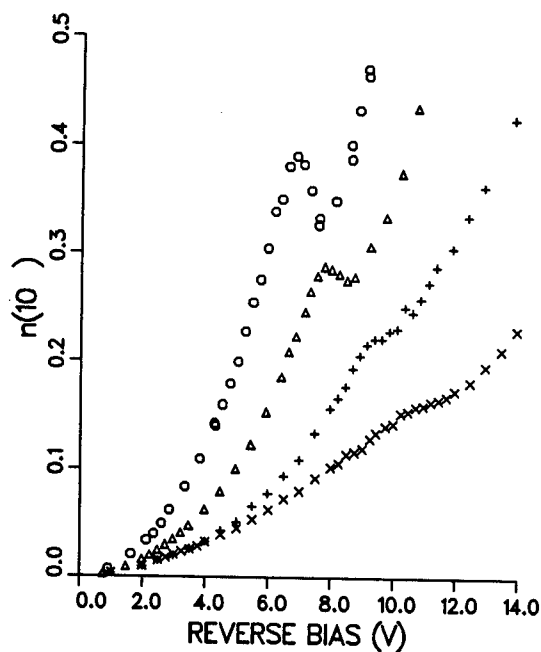
Figure 7. Change in the real part n of the refractive index of the asymmetric coupled quantum-well waveguide as a function of applied bias at four different TM-polarized laser wavelengths: (o) 861 nm; (△) 870 nm; (+) 885 nm; and (×) 900 nm.
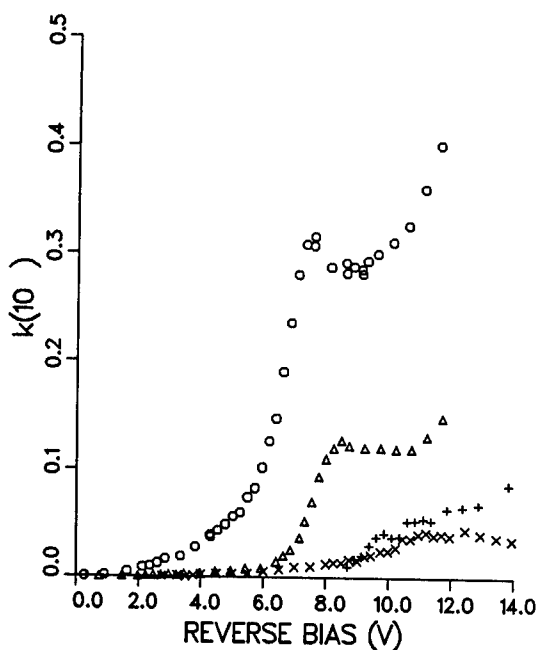
Figure 8. Change in the imaginary part k of the refractive index of the asymmetric coupled quantum-well waveguide as a function of applied bias at four different TM-polarized laser wavelengths: (o) 861 nm; (△) 870 nm; (+) 885 nm; and (×) 900 nm.

The features observed in Fig. 7 have the general "derivative-type" shape expected from the onset of a discrete absorption feature. Such absorptions are apparent in Fig. 8, which shows the bias-induced change in the imaginary part, Δk, of the guide index as a function of applied reverse bias. The electro-absorption in the sample at the shortest two wavelengths clearly shows an excitonic plus band-to-band character, and both the magnitude of the absorption and its dependence on the probe wavelength is consistent with its identification as one of the inter-well transitions described above. Note also that the chirp parameter $\Delta n/\Delta k$ is strongly dependent on wavelength and bias in this structure, and regions favorable for practical application in phase modulation ($\Delta n/\Delta k > 5$) and amplitude modulation ($\Delta n/\Delta k < 2$) are readily found in Figs. 7 and 8. For example, we observed an intensity modulation of 16:1 between 0 and 7.6 V reverse bias at a wavelength of 861 nm, with a corresponding chirp parameter $\Delta n/\Delta k = 1.05$. At 870 nm, we obtain $\Delta n/\Delta k = 18$ between 0 and 4 V reverse bias;

under these conditions, a guide length of 7.2 mm is necessary to obtain a 180-degree phase shift.

Our results on ACQW waveguides have shown that the inter-well transition has a large effect on the below-bandgap optical properties when these structures are biased beyond resonance and, hence, produces electro-absorption and electro-refraction effects that are exploitable in optoelectronic devices.

6. Other Experiments

Recently it has been shown[19] that the extended Bloch states of a superlattice localize when subjected to an external electric field, and that the continuous energy bands of the superlattice evolve into a quasi-discrete spectrum. The energy levels form a Stark ladder, i.e., an equally-spaced energy spectrum given by $E = E_0 + neFd$, where $n = 0, \pm 1, \pm 2, \cdots$, F is the electric field, and d is the period of the structure. Effects associated with this Wannier-Stark localization are of potential significance for devices, since the electro-optic effects associated with localization are very large. Furthermore, intersubband transitions between Stark-localized levels could lead to an electric-field-tunable far-infrared oscillator (Bloch oscillator) with many potential Army applications.

We have discovered a way to utilize the ACQW structure discussed above as a probe for studying Wannier-Stark localization in superlattices. In our novel sample geometry,[6] a pair of wide quantum wells of unequal width, coupled by a thin tunnel barrier, is embedded in the superlattice. (The sample was grown originally to study coupling of hole states[9] in the two large quantum wells at low electric fields, similarly to the electron-state-coupling studies described above.) The period of the superlattice in our sample is considerably shorter than that of similar samples considered by others. Hence, the coupling between the wells in the superlattice is much stronger, and the miniband widths in the conduction and light-hole bands are very large (on the order of a few hundred meV). Also, the width of the heavy-hole band is not negligible, as was the case in previous studies.[19] A diagram of the conduction-band profile in the structure is shown in Fig. 9.

Details of our work on this structure can be found elsewhere.[6] In summary, because of our unique sample geometry, we were able to determine separately the Stark levels of the electron, light-hole, and heavy-hole transitions from our data. Also, our spectroscopic results imply that the coherence length for electrons in this superlattice (which is a measure of imperfections in the sample such as the interface roughness, deviations from periodicity, nonuniformity of the alloy composition in the barriers, etc.)
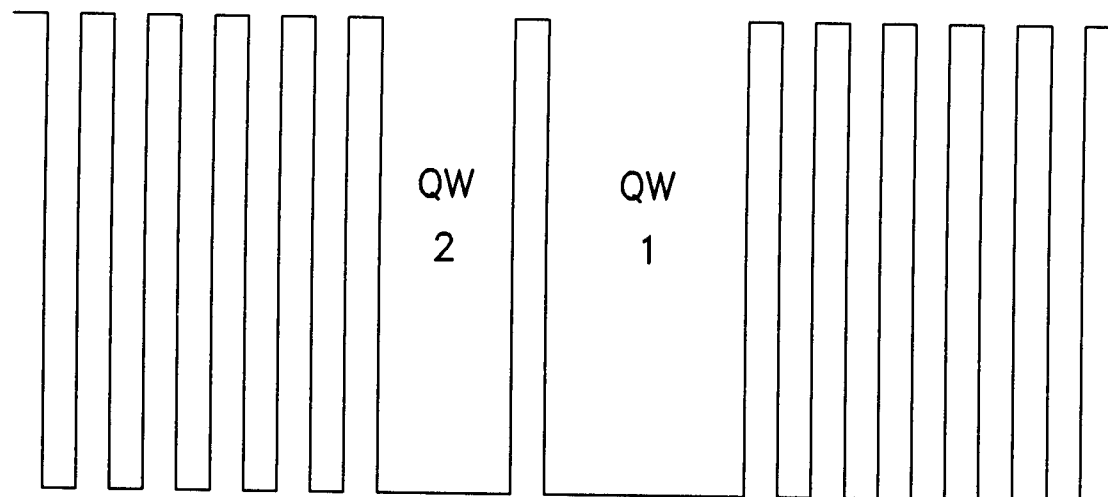
Figure 9. Energy-band diagram of the sample used in the Stark-localization experiments.

is on the order of 1000 Å, which is much larger than that reported by other workers. Another way of stating this result is that the electron wave function is coherent over 50 GaAs/AlGaAs interfaces. Our results have shown that the concept of a Stark ladder is not restricted to systems in which the coupling between adjacent wells in the superlattice is weak. We suggest that Stark ladders may be observable in much shorter-period superlattices than that considered here, i.e., with layer thicknesses approaching atomic dimensions.

In other related work, we have:

- Used PC measurements and interferometer measurements to study coupling of hole levels in samples in which the well ordering is reversed. A number of new physical phenomena were discovered in these experiments.[9]

- Investigated new techniques to grow nearly strain-free, lattice-mismatched ternary and quaternary alloys on GaAs substrates,[20] and used these techniques to grow InGaAs/InAlGaAs ACQW diodes on GaAs substrates.[10] Optical bistability has been observed at room temperature in these structures.

7. Conclusions

A variety of new effects have been observed in asymmetric coupled-quantum-well devices. In many respects, the ACQW system is advantageous for devices, since there are sufficient degrees of freedom to overcome the limitations of the QCSE that occurs in rectangular wells. The size of the effects due to the inter-well transition in these structures does not depend substantially on the width of the larger well, and this parameter may be chosen to give maximum performance of optoelectronic devices *at the wavelengths of interest.* In addition to the width of the widest well and the number of wells, the thickness of the tunnel barrier and the width of the narrower well can be varied to optimize performance. It is expected that, with these additional parameters, devices showing vastly improved performance over those based on the QCSE can be designed to operate in any wavelength region accessible with quantum-well materials.

Acknowledgement

References

1. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, Phys. Rev. Lett. **53**, 2173 (1984).

2. J. W. Little, J. K. Whisnant, R. P. Leavitt, and R. A. Wilson, Appl. Phys. Lett. **51**, 1786 (1987).

3. J. W. Little and R. P. Leavitt, Phys. Rev. B **39**, 1365 (1989).

4. K. W. Steijn, R. P. Leavitt, and J. W. Little, Appl. Phys. Lett. **55**, 383 (1989).

5. R. P. Leavitt, J. W. Little, and S. Horst, Phys. Rev. B **40**, 4183 (1989).

6. R. P. Leavitt and J. W. Little, Phys. Rev. B **41**, 5174 (1990).

7. R. P. Leavitt and J. W. Little, Phys. Rev. B (submitted for publication).

8. K. J. Ritter, R. P. Leavitt, J. W. Little, K. W. Steijn, and S. C. Horst, Appl. Phys. Lett. (manuscript in preparation).

9. J. W. Little and R. P. Leavitt, Phys. Rev. B (manuscript in preparation).

10. R. P. Leavitt and J. W. Little, J. Appl. Phys. (manuscript in preparation).

11. C. Delalande, U. O. Ziemelis, G. Bastard, M. Voos, A. C. Gossard, and W. Wiegmann, Surf. Sci. **142**, 498 (1984).

12. Y. J. Chen, E. S. Koteles, B. S. Elman, and C. A. Armiento, Phys. Rev. B **36**, 4562 (1987).

13. H. Q. Le, J. J. Zayhowski, and W. D. Goodhue, Appl. Phys. Lett. **50**, 1518 (1987).

14. E. J. Austin and M. Jaros, Phys. Rev. B **31**, 5569 (1985).

15. D. F. Nelson, R. C. Miller, and D. A. Kleinman, Phys. Rev. B **35**, 7770 (1987).

16. R. P. Leavitt, Phys. Rev. B (manuscript in preparation).

17. M. Whitehead, G. Parry, K. Woodbridge, P. J. Dobson, and G. Duggan, Appl. Phys. Lett. **52**, 345 (1988).

18. J. S. Weiner, D. A. B. Miller, and D. S. Chemla, Appl. Phys. Lett. **50**, 842 (1987).

19. E. E. Mendez, F. Agulló-Rueda, and J. M. Hong, Phys. Rev. Lett. **60**, 2426 (1988).

20. P. N. Uppal, R. P. Leavitt, S. P. Svensson, J. S. Ahearn, and R. Herring, in J. S. Harris, Ed., *Gallium Arsenide and Related Compounds 1988. Proceedings of the Fifteenth International Symposium*, pp. 191-197.

Novel Magnetic Field Sources for Micro,
MM and Optical Wave Devices (U)

Herbert A. Leupold
U. S. Army Electronics Technology and Devices Laboratory,
Fort Monmouth, New Jersey 07703-5000

I. Introduction:

A perennial barrier to the application of the latest high-powered radiation sources to airborne, ballistic, and the more highly mobile surface vehicles has been the excessive mass, bulk and dependence on power packs of the electron beam focussing magnets that such sources employ. Until relatively recently, attainment of magnetic fields of several thousand gauss over large gaps or volumes depended upon bulky electro-magnets with equally cumbersome power supplies or on large masses of conventional magnet materials whose weight and bulk severely limited application to mobile military devices. Many field configurations were unattainable even with combinations of extraordinarily large mass, high current and small volume.

With the advent of the magnetically rigid high energy product permanent magnet materials, these difficulties became tractable and whole families of previously unattainable devices became viable. Such materials are characterized by very high remanance and coercivity. The former is a measure of a material's ability to provide large amounts of magnetic flux, while the latter is a measure of its rigidity, that is, ability to maintain magnetization in the face of strong opposing magnetic fields that arise from geometric demagnetization effects or from externally applied sources. Such fields are always present in permanent magnet devices, especially in the compact structures that are so important to the military.

Although the high energy product materials (Samarium-cobalt, neodinium-iron-boron, mischmetal etc.) have been commercially available for almost two decades, they have not been exploited to the revolutionary extent warranted by their properties. This seems to be because force

of custom causes designers to employ the new materials to improve performance of old structures rather than to use them in entirely new designs that are not practicable with the old materials. In this regard, the magnetics group at ETDL has been an exception in, that over the past decade, it has striven to formulate general design principles needed to afford efficient exploitation of these materials and to actually apply these principles to obtain efficient, compact and lightweight military devices viz radars, radios, electronic warfare, fuses, magnetic resonance imagers for medical diagnostics, motors, generators and others. This work has resulted in the disclosure of over one hundred patents and the construction of several prototype models of which some of the most advanced will be discussed in this paper. These structures fall roughly into four broad classes with some overlap.

Permanent Magnet Solenoids

Permanent magnet solenoids (PMS), provide uniform fields of
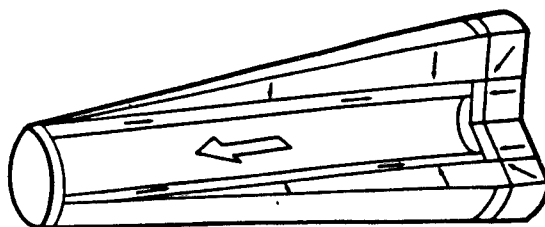
Figure 1. Single chambered Neugebauer structure. Left pole piece is taken as zero potential and outer surface everywhere is lowered to the same potential by inward-pointing cladding magnets. In this way, flux is confined to the interior.
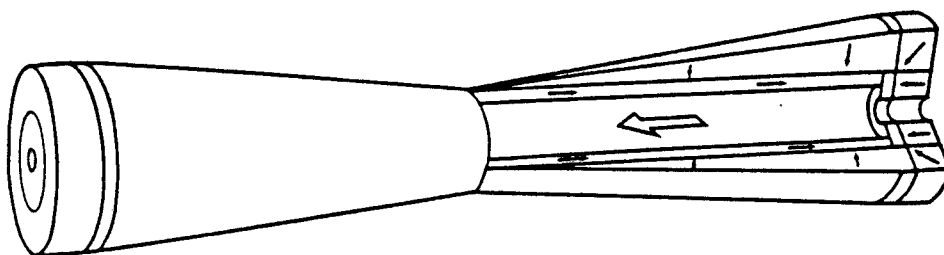
Figure 2. Because the zero potential reference has been moved to the middle from the end, the ETDL structure pictured here has less than half the mass of the Neugebauer design of Figure 1. Solenoidal fields up to 5 kOe are easily produced in this manner, in a structure much lighter than an equivalent electric solenoid and its power supply.

thousands of oersteds over considerable lengths in cylindrical structures. Before the advent of REPM's, such fields were attainable only with electrical solenoids that were generally cumbersome, consumptive of electrical energy and dependent on electric current sources for their operation. The latter are not conveniently portable and do not lend themselves readily to providing field to small spaces with compact structures. In contrast, permanent magnet solenoids generate fields of up to slightly more than half of the remanances of the permanent magnets used (typically about 10 kG for REPM's) in cylindrical spaces of arbitrary dimension without the drawbacks of electric solenoids. The generated magnetic flux is essentially confined to the device and stray fields are minimal. This is accomplished by a configuration that maintains an equipotential everywhere on its outer surface. These configurations consist of three parts: an axially magnetized cylindrical shell which supplies the flux; iron discs that guide that flux into and out of the ends of the cylindrical cavity; and radially magnetized conical cladding which together with axially magnetized discs at the ends, and obliquely magnetized rings in the corners, confines the flux to the interior. Variations of the basic design can produce fields with gradients along the principle axis while maintaining field uniformity over any cross section. Other variations can confine flux to annular ring shaped regions. The PMS are useful in various electron beam devices such as traveling wave tubes, klystrons, cross field amplifiers and gyrotrons. Representative structures are shown in Figures 1-6 in which magnetization vectors are represented by thin arrows and the working field vector by thick arrows.

AXIS OF ROTATION
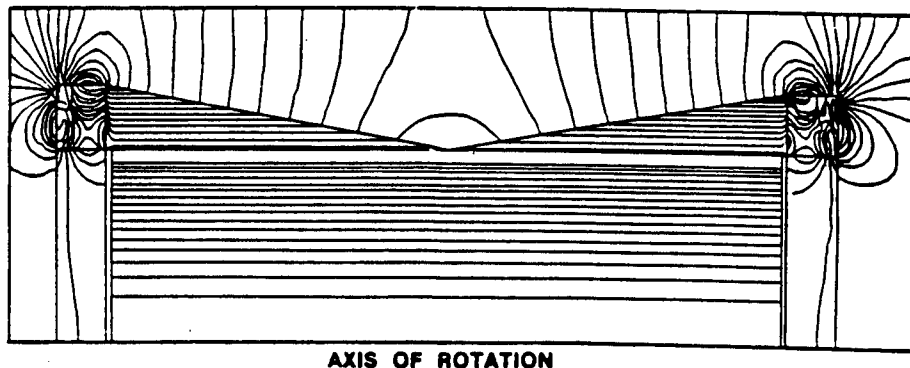
Figure 3. Flux plot of magnetic field produced by the permanent-magnet solenoid of Figure 2. Note the great uniformity over the working space. Apparent flux crowding towards the periphery is because each line represents a unit of flux in an annular ring of given thickness. Actual leakage is only a few percent and is due to imperfect cladding at the corners.

391

**PARAMETRIC STRUCTURE
FOR A LINEARLY VARYING
AXIAL FIELD SOURCE**
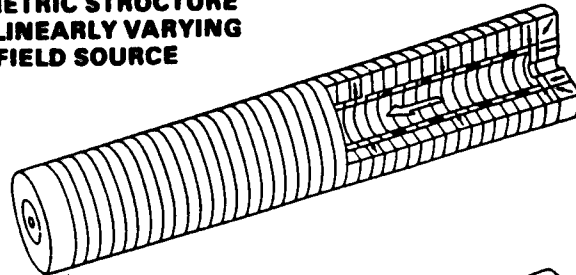
**GEOMETRIC STRUCTURE FOR
A LINEARLY VARYING AXIAL
FIELD SOURCE**
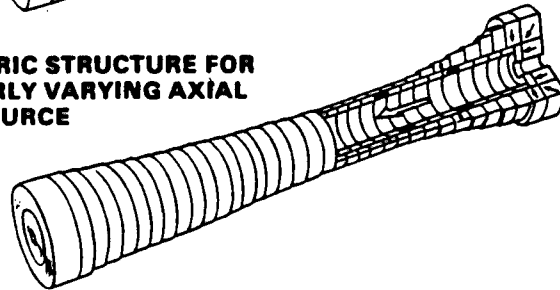
Figure 4. The ETDL permanent-magnet solenoid can be modified to produce fields that vary along the axis. Both of the pictured structures accomplish this. The parametric version controls the variation by modulation of the magnetic properties. In the geometric version, structural dimensions are modulated.
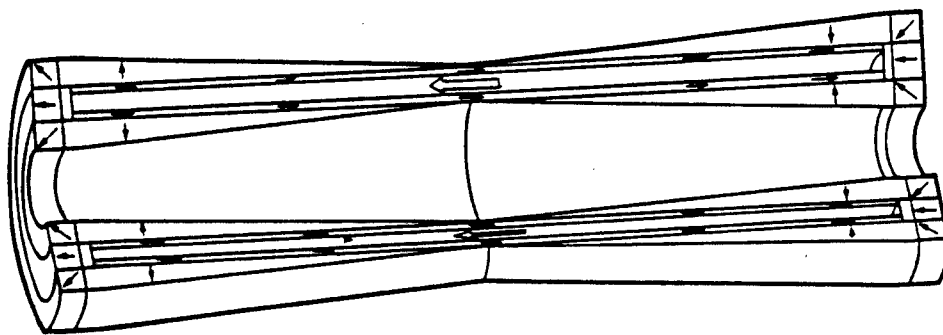
Figure 5. The field in the annular working space is supplied by the axially oriented magnetic shells. The radially oriented shells and end magnets confine the flux. Applications include hollow beam devices.

Figure 6 shows an axially tapered field solenoid that was constructed for an advanced prototype of a mm wave source invented at ETDL. This device requires a solenoidal field that tapers from approximately 1000 oersteds, where the electron beam enters the interaction chamber, to 500 oersteds at the collector end of the chamber. Also, shown is a comparison of the
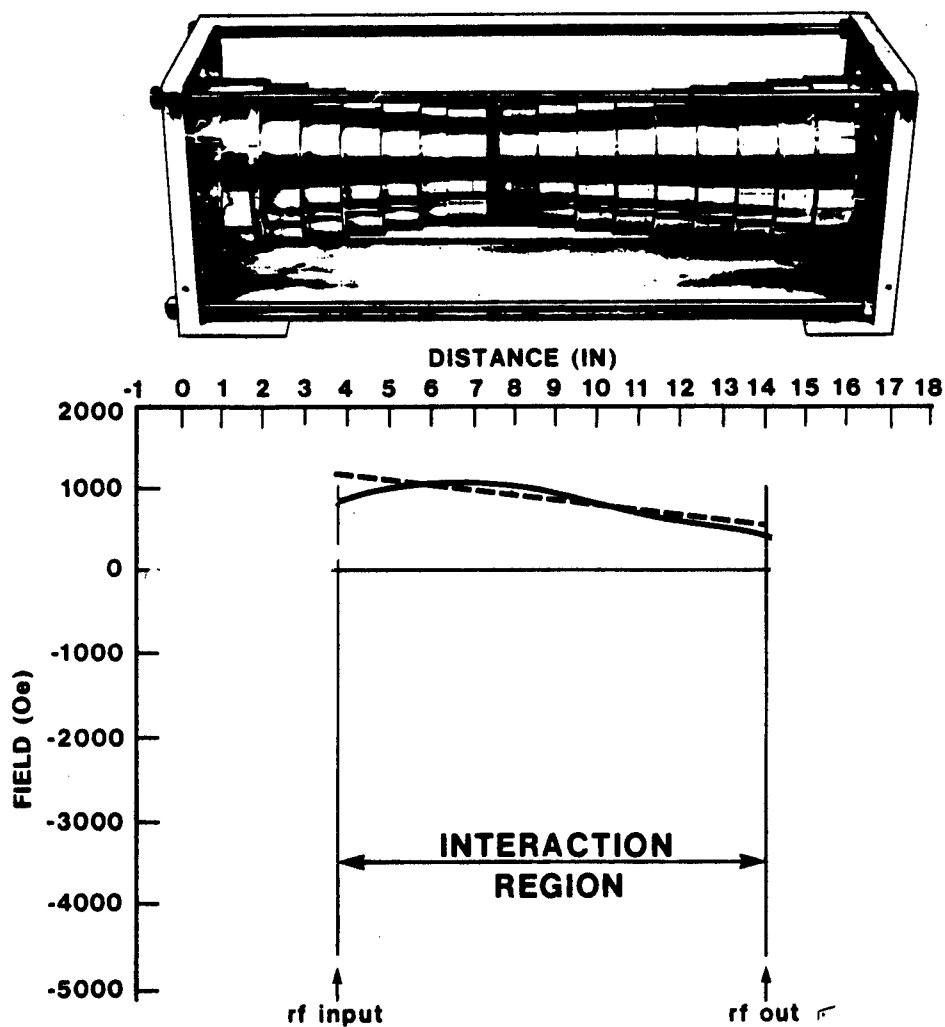
Figure 6. Axially tapered field solenoid.

measured field dependence on axial distance with the calculated dependence. While the compromises and approximations made in construction (e.g. substitution of steps for a continuous taper in the outer magnet) result in some deviation from the desired field, agreement is sufficiently good for operation of the tube. The permanent magnet weighs forty pounds and is to replace an electric solenoid which, together with its power supply, weighs over six hundred pounds. The solenoid is also tied to a current source and consumes considerable electrical energy. Furthermore, its stray magnetic field is strong at a considerable distance from the structure; while for the permanent magnet source, the field is largely confined and therefore affords closer packing of field sensitive instrumentation in its vicinity. The permanent magnet structure clearly affords an enormous enhancement in mobility, efficiency, convenience and tractability so that devices formerly confined to fixed stations, ships and large surface vehicles can now be employed in airborne, ballistic and highly mobile surface devices. Permanent magnet solenoids in various stages of design at ETDL include:

(a). A field source for an advanced, compact, high-power gyro-amplifier for near-mm waves, designed by ETDL at the request of Martin Marietta.

(b). A source for a bi-chambered gyrotron. This source produces a field of 2000 Oersteds in the larger chamber and 500 Oersteds in the smaller. Through techniques developed at ETDL, this can be accomplished in an abrupt step in field at the juncture of the two chambers.
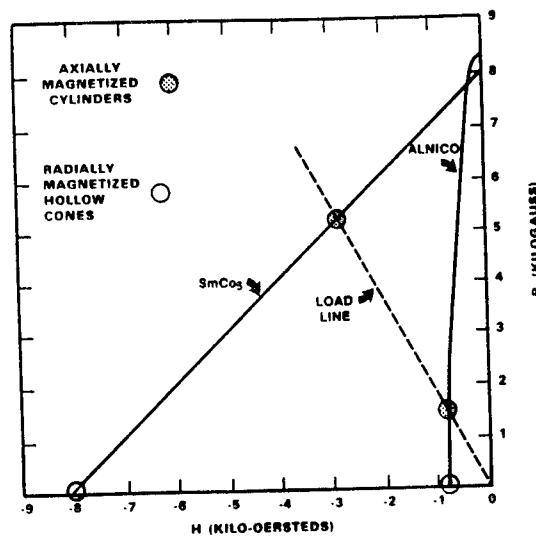
Figure 7. Operating points for permanent magnet solenoid.

(c). A field source for a smoke screen generator to replace a cumbersome, power-consuming electromagnet was designed at the request of Army Ordinance.

(d). A permanent magnet source to lighten cross-field and extended interaction amplifiers.

(e). A klystron magnet for a microwave source for a free electron laser amplifier.

(f). A field source for a satellite-borne crystal grower for NASA.

None of these structures would be viable with conventional magnet materials. This is illustrated in Figure 7. Permanent magnet solenoids all contain magnets operating at point B=0, $H={_B}H_C$ where ${_B}H_C$ is the coercivity. Since the mass and volume of a permanent magnet solenoid are roughly inversely proportional to the square of its coercivity, it is clear that use of Alnico or similar materials would result in a prohibitively bulky structure, two orders of magnitude heavier and larger than Sm-Co magnets producing the same field.

Tranverse Field Sources

Cylindrical structures that produce fields in interior cavities transverse to their axes are illustrated in Figures 8-13. As in the permanent magnet solenoids, the flux in these structures is confined to their interiors. The structures in Figures 9-13 also confine flux to their interiors, but employ no iron pole pieces which help "smooth out" small field distortions engendered by structural defects incurred in the course of manufacture and assembly. They can, however, provide much larger fields than structures with iron poles and are limited in field only by the practical considerations of allowable bulk and weight. Figure 13 shows an adjustable version of Figure 12. Field variation is effected by dividing the basic cylinder into two nested rings, each of which contributes the same field to the interior. When the rings are rotated by the same angle in opposite directions, their vector sum will be in the same direction as the original field but smaller. In this way, any field in the range $+B_{max}$ can be supplied to the interior without the use of electric currents. All the configurations of Figure 8-13 have potential use in magnetic resonance imaging, or in any application in which uniform transverse fields of thousands of gauss are needed. They are of special value in medical diagnostics because of their small bulk relative to superconducting magnets and their freedom from power sources, cryogens and energy expenditure.

At present, these drawbacks limit MRI's to large, wealthy institutions where an entire suite can be dedicated to their use. The structures that employ permanent magnets can be made much more cheaply and afford a degree of portabilty not attainable with superconducting magnets. Therefore, military field use down to divisional or even brigade level is not inconceivable. Moreover, the permanent magnet structures lend themselves readily to miniaturization so that much smaller and more mobile systems could be made for the examination of human extremities and heads. The same miniature systems could also serve as pedagogical devices for the quick training of large numbers of MRI technicians. Applications to anti-terrorist and anti-drug activities also seem feasable; most notably for baggage inspection for contraband at airports and harbors. Several small MRI magnets have already been built and one of half-body size is in the process of assembly and adjustment.
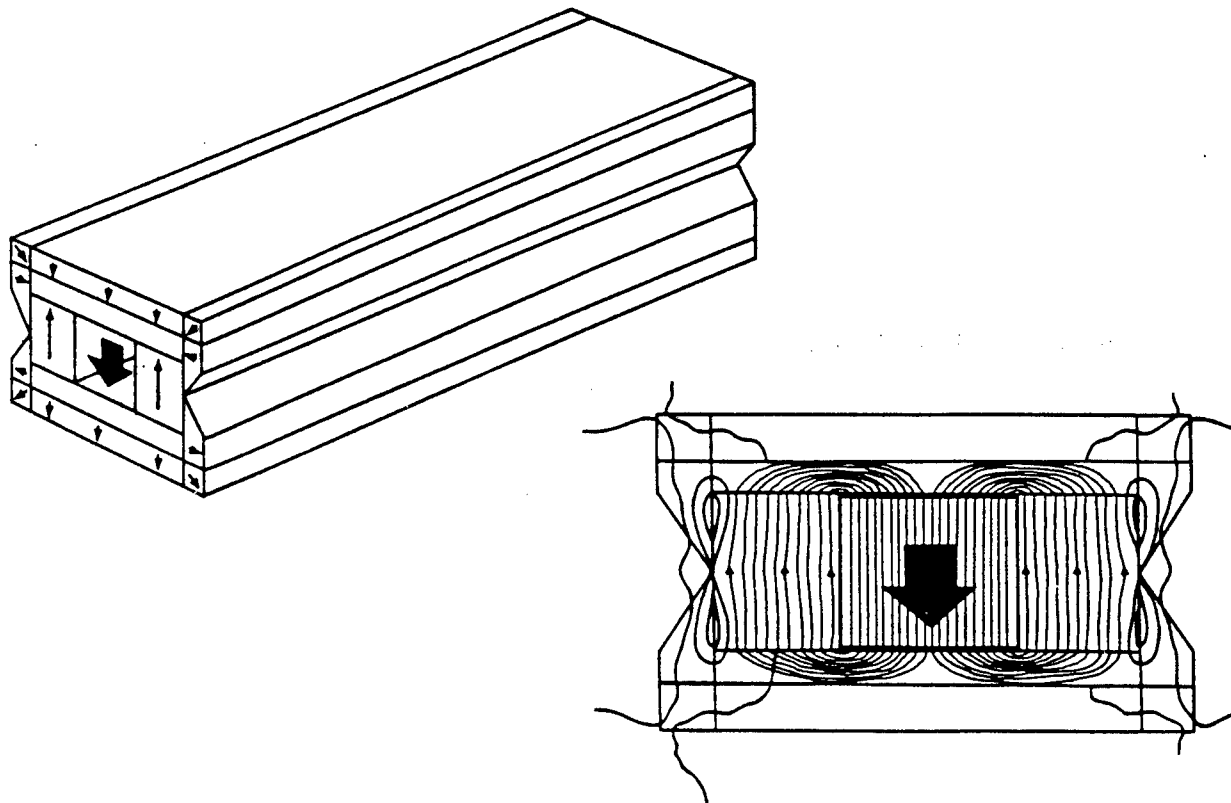
Figure 8. Structure with uniform transverse field in a rectangular working space. The large arrow shows the working field direction and the small arrows show magnet orientations. Possible uses are in NMR imagers and bases for twister structures. Note the field uniformity in the computer flux plot of the cross section.

Figure 9. This structure produces a field of one-half the remanance in a triangular cavity. By successive nesting of many of these structures, arbitrarily high fields are attainable.
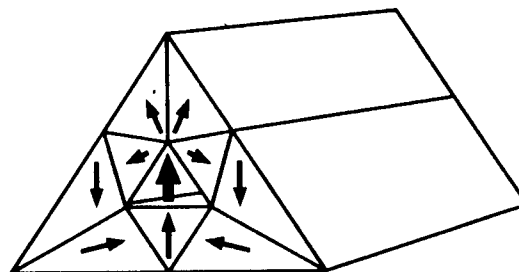
Figure 10. Square permanent magnet structure. As with the rest of the structures on this page, flux is confined to the interior and field augmentaion can be attained by sequential nesting.
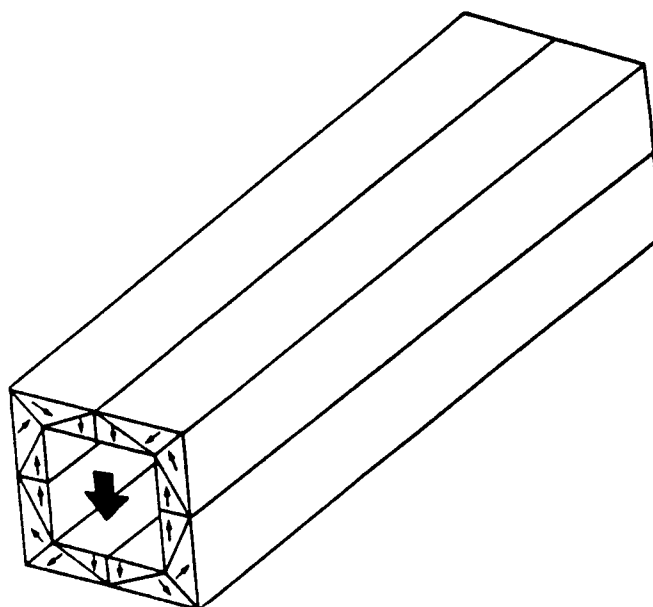
Figure 11. Octagonal
dipole structure.
Fields attained in these
structures are 90 percent
of those of the ideal
circular structure of
Figure 12.
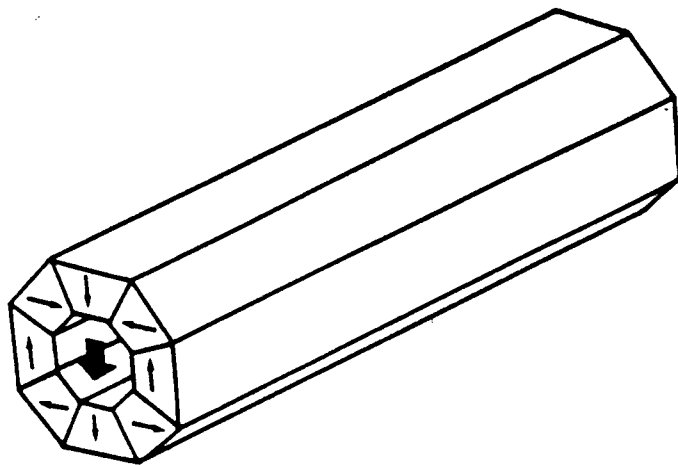
Figure 12. The magnetic
orientation (small arrows)
varies continuously as
$2\theta + \pi/2$. A dipolar field
(large arrow) is thereby
produced in the interior
cylindrical cavity of
magnitude $H=B_r \ln(P_0/P_i)$
P_0 and P_i are the outer
and inner radii of the annular
magnet.
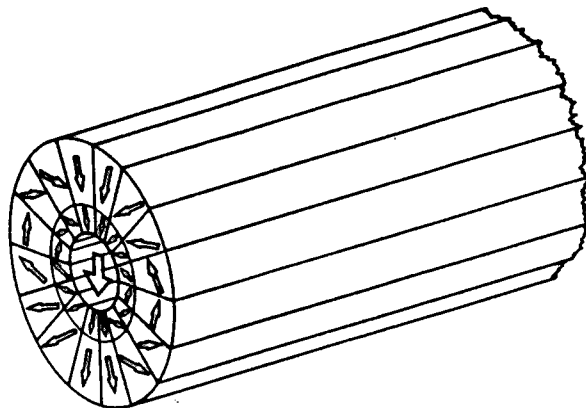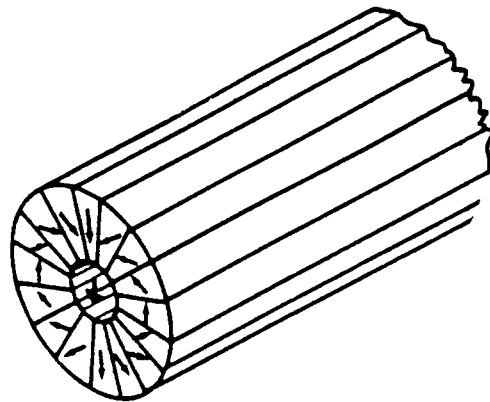
Figure 13. The inner and outer cylindrical shells of the above structure
produce the same field H_0 in the cavity. Therefore, rotation of the
rings with respect to each other yields any field between $+2H_0$ in the
cavity.

Periodic Structures ·

A third set of magnetic configurations is formed from cross sectional slices of the second set arranged in periodic arrays as in free electron lasers (FEL) such as wigglers and twisters (Figures 14 & 15). Such FEL arrays produce higher fields than conventional configurations of similar period, beam diameter and structural mass. They can also be corrected for small field distortions arising from dipoles that must be placed in the inner corners of the polygonal structure to effect the desired compensation. The procedure is particularly valuable in FEL's which are notoriously difficult to adjust. Adjustment is further facilitated by use of the bi-ringed nested structure mentioned above.
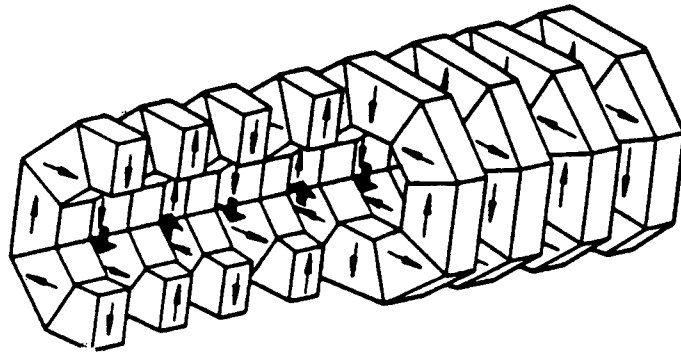
Figure 14. Wiggler structure composed of sectional slices of configuration 11.
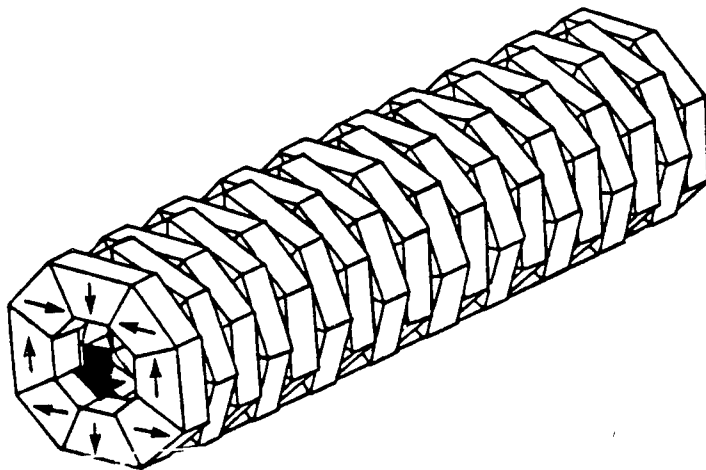
Figure 15. Twister structure composed of sectional slices from the configuration of Figure 11.

In FEL's, an electron beam is sent through a transverse magnetic field that changes directions in a periodic fashion. If the field alternation is roughly sinusoidal as in Figure 14, the electron beam is accelerated from side to side in a direction normal to both its translational motion and the applied field and hence, the device is called a wiggler. The acceleration causes the electrons to radiate energy at the frequency determined by electron velocity and wiggler period. If the proper relationship between velocity, field strength and period exists, the radiation from all parts of the wiggler reinforces and a type of laser action results. When such a relationship exists, the wiggler is called an undulator. If the field remains constant in magnitude but rotates continuously along the structural axis, the electrons are made to follow helical rather than sinusoidal paths and the emitted radiation exhibits circular rather than plane polarization. Such a structure is called a twister or helical free electron laser. Circularly polarized radiation gives radars certain enhanced discriminatory properties compared to plane polarized radiation and is therefore preferable for certain military applications. At the request of the Naval Research Laboratory, a simple permanent magnet twister structure was designed at ETDL that replaces a ponderous electromagnetic field source and its power supply with the usual advantages as is illustrated in Figure 16. A prototype was constructed and found to produce the calculated field of 1200 oersteds as compared with the 500 oersteds delivered by the present coil with the dissipation of 200 amps of electricity.
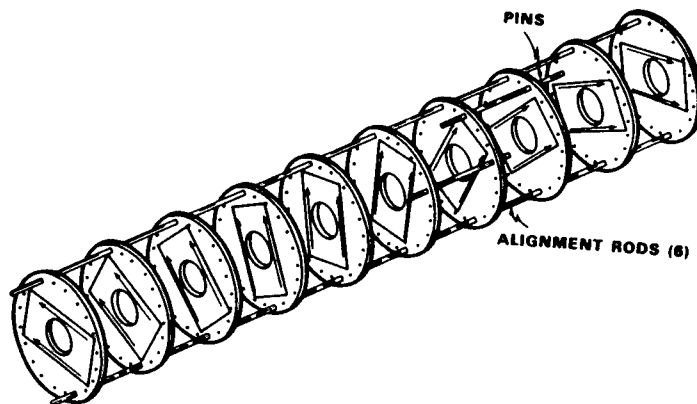
Figure 16. Exploded view of twister structure.

Traveling wave tubes are an important source of mm and microwave radiation for radars and radios. In such tubes, the electron beam is guided by a periodic axial field. Such fields are usually provided by an array of toroidal magnets whose magnetizations are axially oriented in alternate directions and which are interspersed with iron pole pieces that lead magnetic flux into the working space (see Figure 17a). Since such structures are often used in airborne and ballistic devices, where weight and bulk are critical, it is of paramount importance to keep them as light and compact as possible. This is especially true for miniature remotely piloted airborne vehicles where every pound of weight saved can translate into a considerable increase in range or effective payload.

Structures 17b, 17c and 17d were designed and constructed by ETDL to fulfill this need for lighter TWT structures. The configuration of 17b produces the same field as that of 17a with one half as much material while those of 17c and 17d result in from one to two order-of-magnitude mass reductions. The latter two arrays, however, are more difficult to manufacture and adjust and so are not so desirable as 17b except in applications with the most stringent of mass limitations. Structure 17c and 17d are very similar with regard to field to mass ratio and differ mainly in that the focussing ability of 17d is enhanced by a stronger field gradient that exists between its axis and the inner walls of the tube. Both tubes exhibit much better flux confinement than does the conventional structure 17a.

(a)

(b)

(c)

(d)
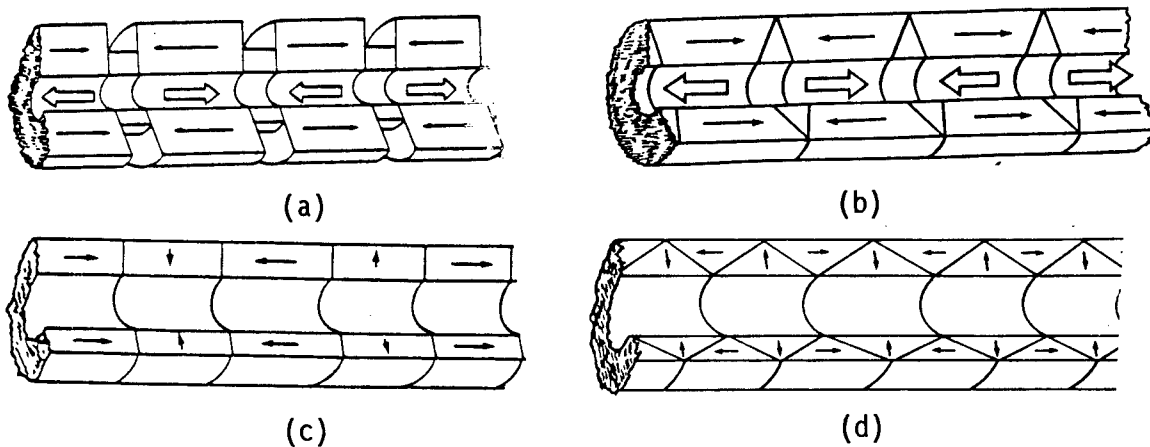
Figure 17. (A) Conventional periodic permanent magnet stack for traveling wave tubes. Iron rings are sandwiched between the axial magnetic guide flux into the bore. (B) A stack with triangular pole pieces. (C) An all permanent-magnet stack consisting of both radially and axially magnetized rings producing the same field as (A) with 1/20th the mass and bulk for a field amplitude of 5.0kOe. (D) A stack with similar performance to that of (C) but with a larger field gradient in the bore.

LEUPOLD

Very High Field Structures

The fourth class of novel structures is generated by rotations of laminar sections of the second class about their polar axes. Figures 9-12 depict representative examples. Such structures produce, in their interior cavities, the very highest fields presently obtainable with permanent magnet structures. For example, in a spherical structure with an inner diameter of one inch, a field of two tesla is obtainable with an outer diameter of 4.5 inches if a remanence of 10 kG is used. With a Nd-Fe-B magnet of 12 kG remanence, the same field is obtainable with an outer diameter of only 3.5 inches.

If a structure of this type is cut in half at the equator and placed on a planar passive ferromagnet such as iron or permendur, the anti-mirror image formed in the ferromagnetic plane produces the same field as the missing half of the original sphere. Such a system is easier to manufacture as it requires only half as many pieces. It also provides more convenient access to the interior as no holes need be bored through the expensive permanent material and different access configurations are easily obtainable in the same hemisphere with different iron plates thereby adding greatly to versatility of use (see Figure 18).
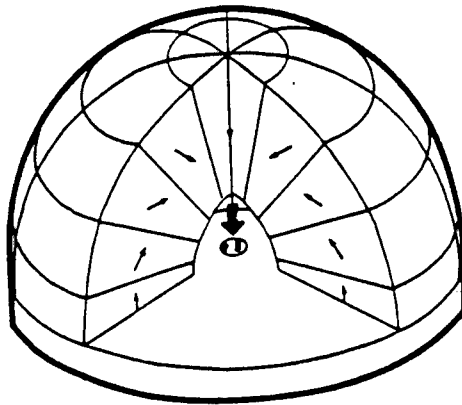
Figure 18. "Magic Igloo", hemispherical magnet set on iron plate.

These structures are useful where very high fields are required such as in some Faraday rotators and short travel beam tubes. Such structures placed in tandem may also be useful as wiggler elements when placed with poles normal to the electon-beam axis or as traveling wave tube elements when arranged in tandem, parallel to the beam axis.

402

Nephila Clavipes Major Ampullate Gland
Silk Proteins: Amino Acid Composition Analysis,
Protein Sequencing, Construction and Sequencing of
Recombinant Genomic and cDNA Libraries (U)

*Stephen J. Lombardi, Mr.
U.S. Army Natick Research, Development
and Engineering Center
Natick, MA 01760-5020

INTRODUCTION

Silks represent an unusual class of fibers generally considered to be protein in composition. Unlike enzymes which are termed globular proteins, silks belong to the structural fibrous protein class which also includes keratin and collagen. Silk proteins exhibit a high degree of crystallinity which is derived from the anti-parallel beta sheet secondary structure. This crystalline array is stabilized by a combination of hydrogen bonding between anti-parallel chains, and hydrophobic interactions between the sheets or layers. These interactions result in a class of fibers with unusual and interesting properties, including high tensile strength and extensibility. The unique properties of silk, and the possibility of genetically engineering controls over fiber structure, greatly expand the availability of biologically-derived materials.

In addition, there is evidence for noncrystalline or amorphous regions in the secondary structure of the silk and it is believed that these regions give rise to extensibility of the fiber and the resulting property of high energy absorption to break. It is the unusual combination of high strength and high extensibility that drives much of the interest in this class of fiber.

In natural systems, the two common sources of silks are the domesticated silkworm, Bombyx mori, and the orb weaving spiders. The silkworm produces one type of silk used in spinning its cocoon during one stage in its life cycle. For domestic silk production the cocoon silk is boiled to remove the soluble sticky sericin protein, and the remaining fibroin portion of the silk is then unwound and used as silk fiber. Orb weaving spiders have the capability to produce many

different silks, each of which is synthesized in a separate set of silk secreting glands in the abdomen. In addition some of the silks, such as the dragline, are produced continuously throughout the lifecycle of the spider. Each of the different silks exhibits different physical properties and functions. These differences are reflected in the amino acid composition of the silk. Some of the silks function in web construction/engineering, in egg cocoon structures, as adhesives, and in prey capture.

There has been some data collected on the physical properties of spider silks. Zemlin [8] and Work and Emerson [7] published data on the amino acid composition and mechanoelastic performance data of different spider silks. Dragline silk from N. clavipes has been reported to have a modulus of 1×10^{10} N/m^2, tensile strength of 1×10^9 N/m^2, and energy absorbed to break value of 1×10^5 J/kg. This compares with the silk worm which exhibits a modulus of 1×10^{10} and a tensile strength of 7.4×10^8. Both fibers exhibit about 18% elongation. Spider silks are of interest because the physical properties of the fibers appear superior to the silkworm silks.

For the silkworm, part of the fibroin gene has been mapped and partial sequencing of the 5' end of the gene was completed [2, 5, 6]. This information, in combination with X-ray data has provided the basic information on the protein structure which indicates discrete crystalline and amorphous regions as reflected in discrete coding regions in the silk gene. Genetic information on the silk worm is extensive because of the commercial interest in this material and in the translational controls over silk expression. No data on the organization of spider silk genes are available.

As in the silk worm, X-ray diffraction data on spider silk implies the presence of crystalline regions dispersed in a matrix of amorphous protein [3]. Additional data on conformation is being developed to further understand the relationships between primary and secondary structures for this class of proteins.

It can be assumed that the mechanoelastic and functional properties of silk fibers correlate closely with their chemical composition and molecular conformation. In addition, the capability of spiders to produce a multitude of silks with very different functions through changes in amino acid composition dictates this system as useful for genetic manipulation for fiber production. The first goal is to clone silk coding genes from the spider into a more useful expression system to increase available amounts of silk. The dragline silk, because of its high tensile strength, was chosen for cloning.

RESULTS AND DISCUSSION

To accomplish the goal of increased silk production the silk gene was cloned. First, genetic libraries were constructed from the spider. High molecular weight genomic DNA was isolated and purified from N. clavipes and then partially digested with restriction enzymes to yield 25 kb fragments. These fragments were cloned into a Lambda phage vector to generate a genomic library. RNA was purified from the major ampullate gland of the spiders. The major ampullate gland is the site of dragline silk production. The mRNA was isolated by density gradient centrifugation and oligo (dT) column chromatography and then reverse transcribed to generate a cDNA library.

To screen these libraries, the native silk protein had to be solubilized, partially hydrolyzed, and sequenced [4]. This was accomplished and the sequence data developed was used to construct DNA probes. The protein composition data confirmed the high percent of short side chain amino acids (glycine, alanine, serine) which permits the close packing density of the beta sheets. The amino acid composition for major ampullate gland silk was determined and glycine, alanine, serine, glutamic acid/glutamine, and arginine were the most abundant amino acids, together comprising over 75% of the total amino acids present. The genomic and cDNA libraries were then screened with these probes. The probes were radiolabelled by the 5' end labelling method and hybridization was determined using autoradiography. Positive clones were then subcloned and expressed in a bacterial host system. Recombinant silk protein was produced and both the clones and the silk are being analyzed. This first phase of work provides the means to produce larger quantities of silk materials for study.

The second phase of the work which is also underway involves selective modification of the natural gene sequence to tailor silk structure and properties to specific functions. To accomplish this goal, molecular modeling studies are being conducted on silk protein sequences to understand the influence of primary sequence on secondary structure. The predicted secondary structure data must then be extrapolated to predicted functional properties of the fibers spun from these sequences. This extrapolation will be validated using protein engineering techniques to enact the sequence changes in the gene, expressing the modified silk proteins, spinning fibers, and then studying fiber properties from the modified proteins.

In general, fiber spinning from recombinant silk proteins will involve mimicking the natural process used by the spider which can be correlated to spinning lyotropic liquid crystals. In the silk gland,

the translated product is present in a metastable state. As the material is processed and then spun at the spinerette, there is a loss of water and the silk protein becomes highly ordered and crystalline. Physical processing appears to effect the major changes in the silk structure and there is no evidence for post-translational chemical modification of the protein. Of interest, based on the amino acid composition data collected, is the fact that spider silk fibers, despite containing a lower percentage of short side chain amino acids, exhibit superior strength properties. This would not be expected and therefore implies a significant role for the processing/spinning steps on resulting fiber properties.

Silk proteins may find application in a number of areas including structural fibers and composite materials. From a composites perspective, a high strength fiber such as silk with its ability to self-assemble into its secondary structure and its reactive functional groups, may provide some unique opportunities in processing and design. In addition, the environmental resistance of natural silks, presumably due to the high degree of crystallinity, would indicate that environmentally stable composite systems incorporating these materials could be considered. Illustrative of the resistance of silk fibers is the fact that proteolytic enzymes are not active in degrading silk proteins. Whether these fibers could also be incorporated into composites for biomedical implants remains to be demonstrated. Work on immobilized enzyme systems using silks for biosensor applications has already been reported [1].

Through genetic engineering we can now consider large scale production of fibrous proteins such as silk. The increased availability of this material will lead to many material applications in the future. In addition, the ability to tailor these structures at the genetic level to meet specific functional requirements further amplifies the potential utility of this approach to materials science.

REFERENCES

1. Demura, M., Asakura, T., Nakamura, E., Tamura, H., J. Biotech. 10, 113-120, 1989.
2. Gage, L. P., Manning, R. F., J. Biol. Chem. 255 (19), 9444-9450, 1980.
3. Gosline, J. M., Danny, M. W. DeMont, M. E. Nature 309, 551-552, 1984.
4. Lombardi, S. J., Kaplan, D. L., J. Arachnol. In Press.
5. Tsujimoto, Y. Suzuki, Y. Cell 16, 425-436, 1979.
6. Tsujimoto, Y., Suzuki, Y. Cell 18, 591-600, 1979.
7. Work, R. W., Emerson, P. D., J. Arachnol. 10, 1-10, 1982.
8. Zemlin, J. C. Technical Report 69-29-CM, AD684333, U. S. Army Natick Laboratories, Natick, MA, 1968.

Calibration of Production Vulnerability/
Lethality Models Based on Live Fire Test Results (U)

Lawrence D. Losie, Mr.
Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

I. INTRODUCTION

In vulnerability and lethality evaluations, we are interested in
estimating expected levels of degradation of vehicular functions for an
encounter between a given munition and a specific armored vehicle. The
term "vulnerability" is used to describe how susceptible an armored
vehicle is to a complete loss for an encounter with a given munition. In
contrast, the term "lethality" is used to describe the same event but from
a totally opposite point of view; i.e., how effective is a given munition
design in destroying a specific armored vehicle.

Thirty years ago, the only way to determine the lethality of a
munition or the vulnerability of an armored vehicle was to conduct a firing
program. Due to the slowness and expense of such an approach, the
Ballistic Research Laboratory (BRL) developed an analytical method to
evaluate the vulnerability of an armored vehicle or the lethality of a
munition design. This method was a compartment level model whose
calculations were based on the analysis of test firings. This model
produces estimates relatively quick. It became our "production" model,
because it was well suited for providing estimates in support of Army
studies. Later, more detailed component level models were developed.
They are more computationally intense than a compartment level model.
Component level models are well suited for certain evaluations, such as
design trade-off studies, where the use of a compartment
level model would not be appropriate. A component level model was used to
provide predictions for Live Fire testing of the Bradley vehicles. This
model was upgraded to reflect results from those tests. It is now used to
calibrate our "production" model for the Bradley Fighting Vehicle Systems.

In this paper, we will present a discussion on our calibration
efforts. Section II briefly discusses our early testing efforts. The
development of computational models in use at BRL will be covered in

Section III.[1] A short discussion on our Bradley experience is presented
in Section IV. Section V discusses different calibration approaches. Some
of our interim results are presented in Section VI. Our concluding remarks
are presented in Section VII.

II. EARLY TEST EXPERIENCE

An extensive collection of firings against tanks was conducted in the
1950s. Tank tests reached their summit in the latter half of that decade
with a set of firings conducted in Canada. These firings are referred to
as the CARDE trials where overmatching shaped charge munitions were
fired against M47 and M48 tanks.[2] The pre-CARDE and CARDE tests were
the largest collection of munition firings against tanks. By the end
of the decade, 1400 shots had been fired. The test results become an
abundant source of information that could be used as a foundation for
developing an analytical vulnerabilty/lethality model.

Associated with the development of a model, a criteria for measuring
effectiveness needed to be precisely defined. In assessing damage
delivered to an armored vehicle by some munition, the chosen criteria must
define what constitutes a "kill" of an armored vehicle. A "kill"
denotes the complete loss of one of the tactical functions of an armored
vehicle. The two tactical functions of interest are the ability to move
effectively and the ability to deliver effective fire. Definitions for
Mobility Kill, Firepower Kill and Catastrophic Kill were initially
established back in the 1950s.[2] Equivalent definitions for each of the
kills are presented below.

   o A Mobility Kill occurs when an armored vehicle is incapable
      of executing controlled movement after being hit and the
      damage is not repairable by the crew on the battlefield.

   o A Firepower Kill occurs when an armored vehicle is incapable
      of delivering controlled fire after being hit and the damage
      is not repairable by the crew on the battlefield.

   o A Catastrophic Kill occurs when an armored vehicle receives
      a hit and the resulting damage is not economically
      repairable.

A Catastrophic Kill generally occurs when the damage from an impact results
in detonation of ammunition or an initiation of a sustained fire. When a
Catastrophic Kill does occur the complete loss of an armored vehicle
ensures that both a Mobility and a Firepower Kill occur.

The key word in the definitions is 'controlled'. Consider a hypothetical case where an armored vehicle in the process of turning receives a hit and the vehicle's suspension becomes jammed which the crew can not fix quickly. The vehicle still moves, but it can only move in a circular or oval path. In such a case, a Mobility Kill would be assessed, because the vehicle's movement is no longer controllable. Associated with the definitions, there is an implicit time requirement for completing battlefield repairs. Expedient repair on the battlefield must be completed within ten minutes or less; if not, a kill would be assessed.

When a munition impacts an armored vehicle, many different damage mechanisms can possibly degrade the functional capabilities of an armored vehicle. There are two primary sources of damage. One is the effect of penetration as the munition encounters components of an armored vehicle along its trajectory. The other is the effect of spall. When a munition perforates the armored shell of a vehicle, the rear face of the armor is ejected as spall debris into the armored vehicle, and this debris has damage potential against a variety of components located in the armored vehicle's interior. There are other sources of damage. Some of these are shock, blast, pyrophorics, vaporifics, hydraulic ram among others. Over a large number of firings their contribution to the total damage is generally minimal; although, in a specific shot they may contribute significantly to the overall damage.

When components receive damage that makes them inoperable, there is a difficulty in comprehending the significance of losing specific components. This difficulty is surmounted by using a device known as a Damage Assessment List (DAL). The first construction of a DAL was done in the late 1950s (circa 1957), and the resulting DAL was known as the Standard Damage Assessment List or SDAL. The actual development of a DAL is accomplished by convening a board of experienced combat Army officers. The board envisions a specific armored vehicle in tactical engagements under a spectrum of all possible combat scenarios covering both offensive and defensive roles. The board considers the question, "If one of the functional systems in an armored vehicle is lost in a hypothetical engagement, how would that loss affect the combat utility of the vehicle?". The result of the board's deliberations is a list of major vehicular systems and for each entry a degradation of combat utility, in terms of Mobility, Firepower and Catastrophic Kills, when such a system is lost.

III. DEVELOPMENT OF MODELS

Using results from the CARDE trials along with some of the test results from pre-CARDE firings, aggregate nonlinear curves, known as

damage correlation curves, were fitted to these data. Damage correlation curves were constructed for a variety of circumstances; i.e., for perforations into the crew compartment, for perforations into the engine compartment, for impacts on external gun, etc. For each circumstance, the corresponding damage curve estimates the expected level of tactical degradation to an armored vehicle due to the cumulative effect of damage mechanisms acting on individual components.

This set of damage correlation curves became an integral part of a computerized methodology which was developed towards the end of the 1950s known as the compartment model or VAMP.[3] In the compartment model an armored vehicle is envisioned as a collection of different compartments. The two primary ones are the crew and engine compartments. Other portions of an armored vehicle can be envisioned as compartments in their own right; e.g., external suspension, main armament, ammunition storage, fuel tanks, etc. Damage assessments are made for every impact location on an armored vehicle. For every encountered compartment at a given impact location, the expected level of degradation is extracted from the corresponding damage correlation curve. The degradation from a damage curve is a function of some explanatory variable which is descriptive of the attacking munition, such as hole diameter or residual penetration capability or penetrator diameter. After evaluating the expected underlying vulnerability, a hit distribution for the attacking munition is superimposed over the armored vehicle. The integration of both of these results in the calculated vulnerability or lethality estimate that is assumed equivalent to probability of kill. In daily use, VAMP provides weighted view average probability of kill estimates for conditions of given a shot (Pk/s) or given a hit (Pk/h).

The compartment model was originally developed for tanks. Subsequently, modifications were implemented that gave us the capability to evaluate munitions against lightly armored vehicles. This computerized code has basically stayed intact over time. Occasionally, a need arises to upgrade VAMP when sufficient data become available. A recent example of this occurred in 1984 when testing indicated such an upgrade was necessary. In this testing, modern shaped charge munitions were fired at improved tank armors. The characteristics of spall debris produced by this interaction were substantially different from debris produced by older munitions perforating steel armor. Since the compartment model is BRL's primary tool for estimating munition lethality, there was a risk that the continuing use of the then existing compartment model could be inappropriate in evaluating munitions against armored vehicles protected with newer armors. Therefore, methods were developed that permitted modifications to the correlation damage curves which reflected the character of the debris produced when modern munitions perforate newer armors.

The compartment model is an expected value model that calculates vulnerability or lethality estimates efficiently and quickly. Since VAMP calculations rely on aggregate damage curves, each of these curves estimates the expected level of degradation when the corresponding compartment is encountered. The estimated damage is independent of the actual location within a compartment, because the estimate represents the damage level averaged over the whole compartment. Due to this, the use of VAMP would be inappropriate for estimating degradation when a given munition is fired at a specific location on a particular armored vehicle. This shortcoming led to the development in the early 1970s of a component level model known as VAST.[4]

In a component level methodology, like VAST, two primary damage mechanisms (i.e., the damage effect of the munition as it impacts the components of an armored vehicle, and the effect of spall debris on interior components when a munition perforates the armor shell) are modeled to estimate the vulnerability or lethality for a munition and armored vehicle encounter. For each impact on a vehicle, this is accomplished by modeling the effect that the damage sources have on each critical component of the armored vehicle. From the calculated component damage, losses to the various systems and subsystems are inferred from the components constituting each of the groups. Then, with the aid of a DAL, the system and subsystem losses are mapped into degradation of tactical vehicular capabilities. This process is repeated for all impact locations on an armored vehicle. As in the compartment model, the distribution of munition impacts can be integrated with the expected underlying vulnerability to obtain weighted view average Pk/s and Pk/h estimates.

Like the compartment model, VAST is an expected value model. Unlike VAMP, it can be used to estimate the level of expected degradation when a given munition is fired at a specific location on a particular armored vehicle. On the other hand, VAST can not produce estimates quickly. Since a component model needs to process and evaluate large quantities of geometric information, VAST is computationally intense.

Later, a slightly less complex methodology called SLAVE was developed.[5] Like VAST, SLAVE is a component level model. Even though SLAVE is less complex than VAST, it is still computationally intense.

IV. BRADLEY EXPERIENCE

Live Fire testing of the Bradley vehicles commenced in 1984 with the Joint Live Fire (JLF) which was a Department of Defense program. Later in that year, the Bradley was pulled from JLF and was placed in the Army's own Live Fire Test (LFT) program. LFT firings against Bradley

vehicles were conducted during the years from 1985 through 1987.[6,7] In
LFT, firings were conducted in two phases. In the first phase, six
munitions were fired at three configurations of the Bradley for a total
of 53 shots. In the second phase, seven munitions were fired at six
different configuations of the Bradley vehicles for a total of 73 shots.

For each firing, BRL used the component code, SLAVE, to calculate the
expected level of tactical degradation. The calculations were made prior
to each shot, and thus they became our predictions. After each shot,
the damage delivered to the tested vehicle was assessed. What was
learned from each firing concerning how damage mechanisms interacted
with the components of the tested vehicles was used to upgrade the
component model.

V. CALIBRATION APPROACHES

As the component model has already been upgraded to reflect the
results observed from the Live Fire test shots, why do we need to
calibrate our compartment model? First, we would like to respond
more quickly in providing estimates to decision makers.   Second, BRL
provides support to Army studies by supplying vulnerability and
lethality estimates for interactions between munitions and armored
vehicles. The studies are computerized war games where the number of
estimates may be few or may be quite large, say 6000 munition/armored
vehicle combinations in a theater level war game. In any year, BRL is
typically tasked to provide estimates for ten or more studies. Thus, for
these reasons, it would be prudent for us to calibrate our compartment
model since the Bradley or derivatives based on the Bradley may be of
interest to decision makers as well as being prominent players in
studies.

To construct compartment correlation curves for the Bradley, three
possible approaches appear likely. One may use results from the Bradley
firings directly or use the component model to generate such curves
or use the component model with the compartment model to adjust the
existing damage curves. Each of these approaches is just as viable as
any other. In this effort we selected the third approach, i.e., use
component and compartment models in conjunction. Reasons for selecting
this approach are discussed below.

In the first possible approach, we could use results from the
Bradley Live Fire tests directly to construct new compartment correlation
curves. We believe such an approach would be inappropriate. Even though
many shots were fired, the number of firings per each tested munition
was too few to give adequate coverage over the whole vehicle. If such
an approach was taken anyway, we believe that the average level of

degradation for each munition would be an imprecise estimator of the true but unknown expected value.

To illustrate this point, consider a hypothetical experiment where we are interested in measuring the average value of some item or thing from observations. The precision of that average is given by its confidence interval. The size of that interval will be proportional to the standard deviation of the population being sampled as well as inversely proportional to the square root of the sample's size. A very precise estimate requires either small variability in a small sample or a large sample size when the variability is not so small.

For our problem, we are most interested where we expect large variability, i.e., for moderate levels of vehicle overmatch. The cases where we expect small variability are concerned with extreme levels of undermatch or overmatch. In the undermatch situation, munitions will rarely perforate and in such cases there is minimal interest. When a munition overmatches a vehicle excessively, the result is obvious – an assassination. So for moderate levels of overmatch, we require large sample sizes per munition to have a precise estimate and thus strong confidence in its intended use. Hence, we cannot use results from the Bradley firings directly and we must resort to an analytical approach. This position is in agreement with the findings of the Board of Army Science and Technology which is an organization of the National Research Council.[8]

In the second possible approach, we could use the component model to generate new compartment correlation curves for the Bradley since observations from the vehicle firings were used to upgrade the component model. For such an approach, the existence of spall models is of crucial importance, because spall debris is one of the primary damage mechanisms. Spall models are empirical models that mathematically characterize the debris formed by perforation of armor. Other tests as part of the Bradley firing program were conducted against Bradley armors to collect the necessary spall debris data, but the tests were only conducted for two shaped charge munitions.

With component model calculations, we can really evaluate only the two munitions for which spall models do exist. With two data points, only linear damage correlation curves can be constructed. A linear model would be difficult to accept, because the analysis of the CARDE firings resulted in correlation curves that were nonlinear. Our only escape from this morass is to consider a spectrum of munitions in a parametric fashion. For these munitions, the debris formed when they perforate armor must be assumed to be identical to the debris produced by one of the two munitions for which a spall model already exists. This solution appears

unacceptable to us, because that assumption would result in compartment curves which we believe would be built upon shaky foundations.

If there is sufficient interest in constructing compartment damage curves by using a component model, the deficiency in spall models needs to be addressed. This lack of spall data to develop debris models has been and still is one of BRL's longstanding needs in conducting component level vulnerability and lethality assessments of armored vehicles.[9] As a partial remedy for this insufficient data base, funding for a future test program would be required. In such a test program, firings of shaped charge munitions against Bradley armors would be conducted to collect necessary debris data. These data need to be collected for a broad range of shaped charge munitions. Analysis of the data would lead to the development of a parametric spall model that could accurately predict the debris characteristics produced by tested as well as untested munitions. With such a spall model, a component model could be used to construct expected value compartment damage curves.

An alternate analytical approach was selected where results from component model calculations are used to modify the damage correlation curves in the compartment model. In this effort, an assumption has been made that the overall shape of the nonlinear damage curves was still valid, but the actual magnitude of the curves may change. We do not intend to calibrate the compartment model with the component model for every possible impact location on the Bradley vehicles. Such an undertaking would be an exercise in futility, because the compartment model uses aggregate damage curves whereas the component model infers damage levels from calculated component losses which are more location sensitive. Rather, after the underlying vulnerability has been integrated with a munition's hit pattern, we would like the weighted view average Pk/h values produced by the compartment model to be able to closely mimic the values produced by the component model.

In this approach, we used shaped charge munitions that were fired against the Bradley vehicles. Vulnerability calculations were made using the component model at 0 degrees elevation (direct fire attack) for each of seven attack azimuths from 0 through 180 degrees, inclusive, at increments of 30 degrees. A hit pattern described by a circular normal distribution with an aim point at the center of the vehicle's presented area in each view and having parametric levels of dispersion was combined with the underlying vulnerability. From this combination, the weighted view average Pk/h values for such encounters were calculated. Likewise, calculations were similarly made using the compartment model. In comparing these initial calculations, significant differences were present in the values produced by the different models.

We believe the differences are primarily due to two reasons. First, damage correlation curves were the result from analysis of the CARDE firings against tanks which were protected by steel armor whereas the Bradley vehicles are protected with aluminum armor. Since spall debris is one of the principle damage mechanisms when armor is perforated, the effectiveness of aluminum debris relative to steel debris in terms of destructive potential against components may be crucial. Secondly, the influence on damage curves that is due to design differences existing between tanks and lightly armored vehicles is unknown. For instance, in tanks the engine is mounted in the rear, but in the Bradley vehicles it is mounted in the front. So connections between driver's controls and the engine compartment are less exposed in the Bradley, and possibly less susceptible to loss than in tanks.

VI. INTERIM RESULTS

In this effort, we set a goal for ourselves to achieve a maximum difference of no more than 0.05 between component and compartment models in their calculated values for the weighted view average Pk/h. If such a goal could be met, we would consider our efforts to be successful. In attempting to reach this goal, the magnitude of the crew and engine damage curves in the compartment model was separately adjusted. In the calculations, they were adjusted for each attack view to minimize differences between weighted view average Pk/h values calculated by the two models over a spectrum of dispersions. This was an iterative procedure that was continued until optimal values for the adjustments could be selected which minimized the maximum differences in magnitude over the largest possible set of views and accuracy levels.

In this forum, we cannot show the optimal levels of adjustment obtained in this approach due to classification concerns. On the other hand, differences between models can be presented. They are shown for one of the tested shaped charges in Tables I and II for Mobility and Firepower Kills, respectively. In the tables, differences between the two models (compartment model - component model) are presented at each attack azimuth and for a weighted average over azimuth. They are presented for varying levels of delivery accuracy corresponding to sigmas from one through ten feet, inclusive, and for a random hit. Differences are generally small over all views and for delivery accuracies having a sigma of three feet or larger. For tighter accuracy levels, the magnitude of the differences is larger but never excessive. At tighter dispersions, we suspect the estimates are becoming aim point sensitive; i.e., only impact locations in near vicinity of the aim point are having strong influence on the calculated estimate.

Differences exhibited in Tables I and II can be envisioned as taking a macro approach in this calibration effort. Details concerning catastrophic vehicular loss permit us to use this macro approach in combination with a micro examination. Examining differences between the two models for Catastrophic Kill, as shown in Table III, reveals that this combination can further improve the level of agreement.

VII. CONCLUSION

In summary, we have discussed several approaches to calibrate our "production" model. The approach that was selected is direct and simple. With the selected approach, magnitude of the damage curves in the compartment model has been adjusted. Adjustments were based upon calculations of the component model that reflect test results from the Bradley firings. This was an iterative process to select optimal adjustment values which minimized the magnitude of the largest differences over the largest selection of views and dispersions. We have presented some of our interim results showing differences in weighted view average Pk/h values between component and compartment models. In this effort, a goal was established to judge for ourselves the magnitude of differences for acceptability. For larger dispersions the goal was easily met, and this goal was generally met with only a few occasional violations at tighter accuracy levels.

At this writing, we are expanding our examination past tactical kills to consider crew and passenger casualties. Towards the end of this year, we expect to have a fully functional "production" model for the Bradley. With this calibrated model, we will be able to respond more quickly to requests for estimates from decision makers as well as providing estimates in support of Army studies.

If our efforts accomplish nothing else, it does show that it is feasible to calibrate our "production" model. Conceptually, an approach similar to the one selected here or some other approach could be used to adjust existing damage curves or to develop entirely new damage curves for any armored vehicle involved in Live Fire testing.
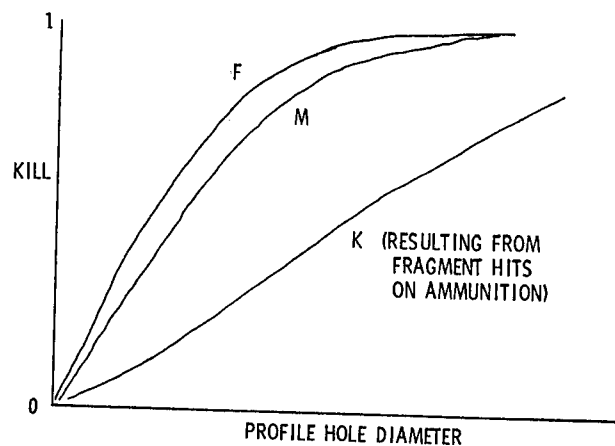
FIGURE 1. EXPECTED KILL FOR CREW COMPARTMENT PERFORATIONS

TABLE I. DIFFERENCES (VAMP-SLAVE) BETWEEN CALCULATED
ESTIMATES OF PK/H FOR MOBILITY KILL

CIRCULAR SIGMAS (FT)	A Z I M U T H S (D E G)							WEIGHTED AVERAGE
	0	30	60	90	120	150	180	
1	.03	-.05	.04	.06	.06	.05	.08	.02
2	.00	.00	.01	.04	.05	.05	.04	.02
3	.00	.00	.00	.01	.02	.02	.01	.00
4	.00	.00	.00	.00	.00	.01	.01	.00
5	.00	.00	.00	.00	-.01	.00	.00	.00
6	.00	.00	.00	-.01	-.01	.00	.00	.00
7	.00	.00	.00	-.01	-.02	.00	.00	.00
8	.01	.00	.00	-.01	-.02	.00	.00	-.01
9	.00	.00	.00	-.02	-.02	.00	-.01	-.01
10	.00	.00	.00	-.02	-.02	.00	-.01	-.01
RANDOM HIT	.00	.00	.00	-.02	-.03	-.01	-.01	-.01

TABLE II. DIFFERENCES (VAMP-SLAVE) BETWEEN CALCULATED
ESTIMATES OF PK/H FOR FIREPOWER KILL

CIRCULAR SIGMAS (FT)	A Z I M U T H S (D E G)							WEIGHTED AVERAGE
	0	30	60	90	120	150	180	
1	-.05	-.04	.01	.05	-.02	-.07	-.07	-.01
2	-.03	.00	.01	.02	.00	-.04	-.04	.00
3	-.01	.01	.01	.01	.01	-.02	-.02	.00
4	.00	.01	.01	.00	.01	-.01	-.01	.00
5	.00	.00	.00	.00	.00	.00	.00	.00
6	.00	.00	.00	.00	.00	.00	.00	.00
7	.00	.00	-.01	-.01	.00	.00	.00	.00
8	.01	.00	-.01	-.01	.00	.00	.00	.00
9	.01	.00	-.01	-.01	.00	.01	.00	.00
10	.01	.00	-.01	-.01	.00	.01	.00	.00
RANDOM HIT	.01	-.01	-.02	-.01	-.01	.01	.01	-.01

TABLE III. DIFFERENCES (VAMP-SLAVE) BETWEEN CALCULATED
ESTIMATES OF PK/H FOR CATASTROPHIC KILL

CIRCULAR SIGMAS (FT)	A Z I M U T H S (D E G)							WEIGHTED AVERAGE
	0	30	60	90	120	150	180	
1	.03	-.07	-.03	-.01	-.04	-.05	.04	-.02
2	.00	.00	.00	.00	.00	.00	.01	.00
3	.00	.00	.00	.00	.00	.00	.00	.00
4	.00	.00	.00	.00	.00	.00	.00	.00
5	.00	.00	.00	.00	.00	.00	.00	.00
6	.00	.00	.00	.00	.00	.00	.00	.00
7	.00	.00	.00	.00	.00	.00	.00	.00
8	.00	.00	.00	.00	.00	.00	.00	.00
9	.00	.00	.00	.00	.00	.00	.00	.00
10	.00	.00	.00	.00	.00	.00	.00	.00
RANDOM HIT	.00	.00	.00	.00	.00	.00	.00	.00

REFERENCES

[1] Due to space constraints, the discussion on testing and development of computational models is limited. A more detailed discussion is presented in P.H. Deitz, A. Ozolins, "Computer Simulations of the Abrams Live-Fire Field Testing," Ballistic Research Laboratory Memorandum Report BRL-MR-3755, May 1989.

[2] Canadian Armament Research and Development Establishment, "Tripartite Anti-Tank Trials and Lethality Evaluation, Part I," November 1959 (UNCLASSIFIED).

[3] C.L. Nail, E. Jackson, T.E. Beardon, "Vulnerability Analysis Methodology Program (VAMP): A Combined Compartment-Kill Vulnerability Model," Computer Sciences Corporation Technical Manual CSC TR-79-5585, October 1979.

[4] C.L. Nail, "Vulnerability Analysis for Surface Targets (VAST) - An Internal Point-Burst Vulnerability Assessment Model - Revision I," Computer Sciences Corporation Technical Manual CSC TR-82-5740, August 1982.

[5] D.A. Ringers, F.T. Brown, "SLAVE (Simple Lethality and Vulnerability Estimator) Analyst's Guide," Ballistic Research Laboratory Technical Report ARBRL-TR-02333, June 1981.

[6] D.C. Bely, G.L.Holloway, "Bradley Survivability Enhancement Program - Phase I Results - Volume I - Results Summary and Analysis," Ballistic Research Laboratory Special Publication SP-49, November 1985 (SECRET).

[7] U.S. Army TECOM, BRL, AMSAA, CSTA and OSG, "Bradley Survivability Enhancement Program, Volume I, Phase II Live Fire Test Report," June 1987 (SECRET).

[8] Board on Army Science and Technology, "Armored Combat Vehicle Vulnerability to Anti-Armor Weapons: A Review of the Army's Assessment Methodology," National Academy Press, 1989.

[9] D.L. Rigotti, P.H. Deitz, D.F. Haskell, M.W. Starks, D.P. Kirk, O.T. Johnson, J.R. Jacobson, W. Kokinakis, J.T. Klopcic, G.A. Bowers, "Vulnerability/Lethality Assessment Capabilities - Status, Needs, Remedies," Ballistic Research Laboratory Special Publication SP-74, December 1988.

Workload, Target Acquisition and Piloting Performance:
Psychological and Physiological Predictors (U)

*Jeffrey H. Lukas, Dr., Richard A. Monty, Dr., Mary E. Dominessy, Ms.,
Frank J. Malkin, Mr., and Lynn C. Oatman, Dr.
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground, MD 21005-5001

Helicopter pilots must fly, navigate, and acquire targets while monitoring complex visual displays. In addition to their traditional role, Army pilots in the future may be required to support air defense ground forces in locating and defeating enemy aircraft. Helicopters assigned this enhanced mission capability will require a situation display[1] to present the location of friendly, enemy, and unknown fixed and rotary wing aircraft. In this high workload environment, the situation display must be designed optimally to provide information the crew needs to acquire targets without overloading their processing capacity or adversely affecting their flying. The Human Engineering Laboratory (HEL) began a research program examining numeric, verbal, or graphic display formats under several levels of workload while piloting a helicopter simulator to determine which format met these criteria. In addition, the study of display format and piloting performance[2] afforded an opportunity to advance research aimed at developing techniques for selecting individuals for high workload, stressful positions. The Army needs to select and train the best candidate pilots who possess all the mental, emotional, and physical qualifications to succeed. A crucial consideration must be their ability to function effectively with high mental workloads and the resulting stress it produces. As part of a research program examining cortical functioning, personality, and stress susceptibility,[3,4,5] an evoked potential (EP) procedure was developed which predicted performance in a mental workload task.[6,7] EPs are a particularly promising screening device since they cannot be faked or altered and provide a direct, objective assessment of the processing capacity of the brain. The present study was undertaken with two main objectives: first, to determine if the EP procedure would predict performance during an operationally realistic task of flying a helicopter simulator while acquiring targets; and second, to compare current Army pilots' EP, personality, and workload performance data with previous data collected from college students and nonpilot Army personnel. The goal of

this basic research program was to develop a screening device to select pilots, air traffic controllers, and other personnel who must perform flawlessly under difficult and demanding conditions.

The study of cortical functioning is based on the EP recorded over the vertex scalp to brief, repetitive light flashes that vary in brightness.[8] Augmenters produce larger EPs as the light flashes become brighter. Curiously, reducers have the opposite and surprising response of smaller EPs as the intensity increases, indicating a state of reduced cortical functioning.[4] Augmenters and reducers differ in a number of dimensions including personality, emotional responsiveness, and performance under high mental workloads. Human augmenters are sensation seekers[3,9] and appear to relish higher levels of stimulation through experiencing new, exciting, and sometimes dangerous activities. Augmenter animals perform better than reducers during aversive noise stimulation[5] and during several food-rewarded tasks.[10] Human augmenters were found to perform significantly better under high mental workloads in a keeping track task.[6,7] Results from this line of research indicate that augmenters cope better, seek higher levels of stimulation, and perform better under high mental workloads, aversive stimulation, and mild stress. The EP cortical assessment procedure appeared promising as a potential screening device.

It was necessary to transition these basic research findings to a more applied arena to determine if the EP procedure predicted proficiency on an operationally realistic Army task. In the study of situation display formats,[2] would augmenters perform better than reducers at the dual task of acquiring targets and flying the helicopter simulator? More importantly, since subjects for this study were current National Guard pilots who had passed through selection and training and functioned daily under high workloads, theory predicted they would be augmenters. Therefore, it would be very informative to compare pilots with nonpilots from previous studies in terms of augmenting-reducing, personality, and workload performance to further develop a screening device to select pilots, air traffic controllers, and other personnel who must function under high workloads and stress. Data are presented examining four main hypotheses: first, pilots are more sensation seeking, more extraverted, and less neurotic than nonpilots; second, pilots perform better under high workloads in the keeping track task; third, augmenters will be far more prevalent among pilots than nonpilots; and fourth, augmenter pilots will perform better than reducers under high workloads in the keeping track task and helicopter simulator.

Subjects

Eighteen male helicopter pilots from the Delaware and Maryland National Guard units volunteered to participate in this study. Performance data were lost for two pilots on the helicopter simulator as a consequence of equipment failure. The pilots averaged 1204 flight hours in rotary wing aircraft with a range of 250 to 4500 hours. All had normal vision and were not receiving medication at the time of testing. Their average age was 33.8 years, with a range of 24 to 43 years.

Apparatus

Light flashes were presented in Maxwellian-view to the right eye. A series of lenses collimated the light emanating from a 150-W tungsten-halogen lamp and focused the beam to a 2-mm image at the pupil. With this system, shifts in pupillary diameter with intensity could not alter the amount of light reaching the retina. White light flashes were presented at 1/sec with a 100-msec duration. Infrared rays were removed with a filter. The light subtended a 22° visual angle and measured 5.9×10^5 trolands at the brightest intensity. Light intensity covered a 3 log range from this maximum intensity in seven 0.5-log steps using neutral density filters. The lamp, filters, and lenses were outside a double-walled, acoustically shielded chamber. Subjects sat in the chamber on a hydraulic chair which could be adjusted along with a chin and forehead rest to bring them into alignment with the light flashes. Subjects fixated on a cross hair in the light beam to maintain alignment. A video camera, microphone, and speaker were located in front of the subject in order to ensure that the subject was ready for each trial and in good alignment.

Target acquisition and piloting performance data were collected using a fixed base helicopter simulator based on UH-60 Blackhawk flight parameters with a side arm controller rather than a center stick for controlling roll and pitch. A computer generated "out-of-the-window" scene consisting of a river wandering among trees, hills, roads, and buildings was presented on a 1.57 m high X 2.06 m wide viewing screen located approximately 2.13 m from the subject. All system, navigational, and flight data (i.e., speed, altitude, heading, etc.) were presented on video displays located on the instrumentation panel. The hook commands and situational displays were presented on a 15.2 cm X 20.3 cm CRT display located in the lower center of the instrument panel. Subjective workload measures were obtained using the NASA Task Load Index.

Procedures

Two separate studies were conducted with eight pilots participating first in the EP, personality, and workload study in the morning and then completing the helicopter simulator and target acquisition study in the afternoon. The remaining pilots participated in the opposite order.

Recording electrodes were placed on the scalp over the occipital, parietal, and vertex cortical areas and on the zygomatic bone beneath the right eye for recording the electroretinogram (ERG). A reference electrode was placed on the right earlobe, while the left ear served as a ground. Electrodes were reapplied if impedance exceeded 3 Kohms. Following electrode application, subjects were taken to an electrically shielded, sound-attenuated chamber where they were seated on a hydraulic chair and positioned to allow the light beam to pass through the pupil of the right eye. The left eye was covered with an opaque patch. Subjects were instructed to relax and maintain fixation on cross hairs in the center of the light flash. Subjects were shown how to maintain proper alignment and fixation. They were then adapted for 5 minutes to an ambient illumination of 34 ft L. Seven flash intensities were presented in steps of 0.5-log intensity in ascending order, with 100 light flashes presented at each intensity. Following a 4 minute break, the entire series of seven intensities was repeated. Subjects counted the light flashes and reported this number at the end of each trial. The actual number of stimuli presented on each trial varied with the number of artifacts rejected by the computer. Subjects verbally initiated each trial when in proper alignment and ready to concentrate, and were given rest breaks throughout the experiment. Cortical activity was amplified 10,000 times with a bandpass of 1-100 Hz. The ERG had a bandpass of 1-300 Hz and was amplified 5000 times. The EPs were computer averaged using 1024 data points and a 300-msec sweep. Sweeps were rejected if they exceeded 90% of full-scale voltage (50 uV) to eliminate data contaminated with eye blinks or other artifacts. Artifacts were relatively homogeneously distributed across intensity with 82% of the trials having no artifacts and only 3% of the trials having more than five artifacts.

Following the evoked potential procedures, subjects completed two personality inventories: Eysenck's Personality Questionnaire - Revised (EPQ-R) short scale which is a 48-item questionnaire containing four scales of 12 items each: (1) Psychoticism, (2) Extraversion, (3) Neuroticism, and (4) Lie. The second questionnaire was Form V of Zuckerman's Sensation Seeking Scale (SSS) comprised of four scales of 10 items each and a total score. The four scales are: (1) Thrill and Adventure Seeking, (2) Experience Seeking, (3) Disinhibition, and (4) Boredom Susceptibility.

Following a short break, subjects began a keeping track task. Subjects viewed a small screen (30 X 30 cm) in which four letters (P, Q, R, S) were presented in a random order with a duration of 100 msec and an interstimulus interval of 1.4 sec. Each trial consisted of 15 letters. Subjects' task was to keep a separate mental count of each letter and report the number at the end of the trial. The keeping track task was designed to manipulate mental workload by increasing the task difficulty from keeping track of one letter to keeping track of all four letters simultaneously. Subjects were given three training trials at each workload level for a total of 12 training trials. They then received nine trials at each workload level in increasing difficulty and were provided verbal feedback following each trial. Keeping track of four letters, the highest workload task, was repeated an additional nine trials. At the end of the study, subjects rated how difficult and interesting each task was and provided information about their strategy for keeping track.

Target Acquisition Procedure

Using a CRT display in the fixed base simulator, subjects were given mission instructions to search for a particular friendly, hostile, or unknown aircraft on a situation display. Three different instructional formats were contrasted, specifically, a text mode in which the aircraft or target to be "hooked" was identified textually (i.e., "Hook Rotary Wing Hostile 49"); a graphic mode in which targets were identified symbolically with triangles, squares, and diamonds; and a numeric mode in which only the track number of the target was presented. Following acknowledgement of receipt of the instructions, the screen blanked and was replaced with a situation display showing the designated aircraft along with 11 others. The subject had to search a CRT situation display and touch ("hook") the designated target as quickly as possible. Subjects were required to perform this task as a copilot (low workload) and while piloting the simulator (high workload condition).

Evoked Potential Analysis

Each pilot's visual evoked potentials were collected for two series of seven flash intensities and summed to produce one EP for each intensity. The vertex P1-N1 component[8] was measured using a computer cursor routine. If P1 did not occur at the appropriate latency (90 to 100 msec) or the VEP was excessively noisy, then the numeric average of the two VEPs produced by the surrounding intensities was substituted in its place. Two measures of cortical functioning were computed for each pilot: the slope of the linear regression line for the vertex P1-N1

amplitude for all seven intensities and the slope for the brightest five intensities, both expressed as uV amplitude per log intensity (uV/log).

Helicopter Simulator Performance Measures

This study consisted of a 2 X 2 X 3 factorial design with Crew Status (performing the target acquisition task as pilot or copilot), Task Difficulty (two or four digit track numbers), and Presentation Format (text, graphic, or numeric) all as within subjects variables. Subjects received 12 trials at each of the 12 treatments in a counterbalanced order and completed the NASA subjective workload scale after each treatment. Three measures of target acquisition performance were analyzed: response time 1, the time required to acknowledge the hook command; response time 2, the time to hook the target once the situation display was presented; and total response time, the sum of RT1 + RT2. No analysis of errors was undertaken since the error rate for the target acquisition task was low (3.5%) and errors were randomly distributed across treatments. Two additional performance measures were collected while the subjects were piloting the simulator, namely, RMS error airspeed and altitude measured as deviations from the required 50 knots and 150 feet altitude above the river.

Results

Sample size for all analyses is 18 except for the helicopter simulator task, in which data were recorded for 16 subjects. For all statistical tests, a probability level of $p < .05$ is considered significant. Results are presented in regards to four hypotheses: (1) personality -- pilots are more sensation seeking, more extraverted, and less neurotic than nonpilots; (2) keeping track performance -- pilots perform better under high mental workloads than nonpilots; (3) cortical assessment -- pilots are more likely than nonpilots to be augmenters; and, (4) augmenter pilots will perform better in the keeping track task as well as in the dual task of target acquisition and piloting than reducer pilots.

Personality

Pilots' personality scores on the Sensation Seeking Scale and Eysenck's Personality Inventory were compared with two groups of male nonpilots (N=20 each) comprised of military and college students who participated in previous studies examining personality correlates of augmenting-reducing. Separate one-way analyses of variance were computed for each personality score comparing the three groups. As shown in Table 1, only the Neuroticism scores differed among the three groups, $F(2,55) = 9.43$, $p < .001$. The pilots scored significantly lower in neuroticism than

either group of nonpilots. The pilots' extraversion scores did not differ from the nonpilots' scores; however, both pilots and nonpilots scored considerably higher in extraversion than normative samples.[12] Scores on the SSS did not differ for pilots and nonpilots and were similar to published norms.[13]

Table 1

Means and standard deviations for the EPQ-R and the SSS for pilots and for two samples of male, nonpilots. Also presented are the norms and analyses of variance F scores comparing the pilots and nonpilots.

	Pilots		Nonpilots I		Nonpilots II		Norms	
	N=18		N=20		N=20			
Scales	Mean	SD	Mean	SD	Mean	SD	Mean	F
EPQ-R								
E	9.2	2.7	9.0	3.2	8.4	2.9	6.91	0.39
N	1.6	1.3	4.9	3.1	4.9	3.1	5.46	9.43***
P	2.3	1.5	2.5	1.8	3.0	2.1	3.20	0.75
L	4.3	3.0	3.4	2.5	4.3	2.2	3.25	0.80
SSS	19.8	5.9	20.9	6.6	20.7	5.6	21.20	0.17

*** $p<.001$

Keeping Track Performance

Pilots' (N=18) workload performance was compared to data previously collected from volunteer laboratory scientists and engineers (N=10) performing the identical keeping track task. Subjects' total errors were collapsed across nine trials under each of the five workload levels: keeping track of one, two, three, or four letters simultaneously, with keeping track of four letters repeated an additional nine trials to give the fifth level. A MANOVA was performed on the transformed error scores, log (errors + 1.0), with Groups (Pilots versus Nonpilots) as a between variable and Workload (five levels) as a within variable. The main

effect for Groups was significant, $F(1,26)= 9.3$, $p<.005$, with the pilots averaging 48% more errors than the scientists and engineers (44.0 versus 29.8 average errors). Errors increased for both Groups as mental workload increased, $F(4,23)= 193.0$, $p<.0001$, and the Group X Workload interaction was not significant.

Cortical Assessment

The slope of the linear regression relating the amplitude of the vertex P1-N1 component for the seven flash intensities was calculated for each of the 18 subjects and expressed as microvolts per log intensity (uV/log). Table 2 compares the pilots' slope data with slopes previously collected from 19 male nonpilot college students using the same evoked potential procedure. As predicted, the pilots as a group were more augmenting than the nonpilots, with slopes of .73 versus .23 uV/log. A median test comparing the slopes for the two groups indicated a significant difference, $X^2=4.56$, $p<.02$. The range of augmenting slopes was very similar for the pilots and nonpilots. However, both the incidence and range for reducing slopes were quite different, with eight nonpilots (42%) having negative slopes ranging from -.04 to -.82 uV/log, while one pilot (6%) had a marginally negative slope of -.08 uV/log.

Table 2

Comparison of vertex amplitude slopes for pilots and nonpilot, male college students.

Group	N	Slope uV/log	Minimum Slope	Maximum Slope	Reducers	Percent Reducers
Pilots	18	.73	-.08	2.06	1	6%
Nonpilots	19	.23	-.82	1.65	8	42%

Augmenting-Reducing and Workload Performance

In previous studies, differences in keeping track performance were seen as a significant Groups (augmenter-reducer) X Workload performance interaction.[7,11] Both groups committed few errors while keeping track of one or two letters simultaneously. At high workloads, keeping track of three or four letters simultaneously, the augmenters performed significantly better than reducers with approximately one-half the

errors. A MANOVA with Groups as a variable was clearly impossible in the present case with only one pilot in the reducer group. Therefore, in this primarily augmenter sample, the analyses focused on the correlations among the evoked potential slope and workload performance measures. A negative correlation was predicted given the hypothesis that the more augmenting the pilot (i.e., the greater the magnitude of the positive slope) the fewer errors he should commit. Table 3 presents these correlations with the slope calculated from all intensities as well as the slope to the five brightest intensities. One negative correlation was significant, and 10 of 12 correlations were negative providing some support for the hypothesis.

Table 3

Correlation of slope for all seven intensities and high slope for five brightest intensities with errors on keeping track of one, two, three, four (a,b) letters or total errors.

	1	2	3	4a	4b	Total
Slope	-.48*	-.34	.07	-.39	-.04	-.28
High Slope	-.23	-.21	.11	-.22	-.12	-.19

$*p<.05$

A primary purpose of this research was to determine whether the cortical measurement procedure would predict performance on an operationally realistic helicopter simulator in which mental workload was experimentally manipulated. Again, negative correlations were predicted with the more augmenting pilots having shorter response times and lower altitude and airspeed RMS errors. However, examination of the correlation matrix indicated there were no significant correlations with either the slope to all intensities or the slope to the brightest intensities.

Personality scores and keeping track performance were also correlated with helicopter performance. Extraverts tended to have better piloting performance with the correlation with RMS error airspeed significant ($r= -.67$, $p<.01$) and a marginally significant correlation with average performance across all performance measures in the helicopter simulator ($r= -.49$, $p<.06$). In addition, performance on the

second set of keeping track of four letters correlated positively with all helicopter performance measures and was significant with average performance across all performance measures in the helicopter simulator (r= .61, p<.02). Pilots making a lot of mistakes in the keeping track task tended to have longer target acquisition response times and worse helicopter performance.

Conclusions

The most important finding to emerge from this study was that the pilots as a group were more augmenting than nonpilots. In fact, 18 of 19 (94%) pilots were augmenters and only one pilot was a reducer with a marginally negative slope of -0.08 uV/log. In previous studies using visual stimuli, approximately 40% of the subjects were reducers,[3,7,14] a much higher percentage than found in the pilot group. Although caution should be exercised in directly comparing results from different studies, the magnitude and consistency of the difference across studies indicates the finding is robust. Reducing was not only less prevalent among the pilots, but also did not fit the usual reducing pattern. Typically, most reducers have small vertex potentials that decrease in amplitude as flash intensity increases (type 1). The reducer pilot had larger than average EPs for all intensities, and especially for the two dimmest intensities where the EPs were almost three standard deviations larger than the mean (type 2). A type 2 reducer may indicate an individual who is more responsive or sensitive than the type 1 reducer to low or moderate intensities. Although the sample sizes are too small for statistical comparisons, it is interesting to note that type 2 reducers occurred in previous studies and typically performed well in the keeping track task unlike the type 1 reducers who often made many errors. The tentative conclusion is that type 2 reducing indicates enhanced responsiveness to moderate levels of stimulation rather than reduced cortical processing capacity under high levels of stimulation. Therefore, an essential refinement of the cortical assessment procedure entails quantification of the form of the EP amplitude--stimulus intensity function in addition to measuring the linear slope.

Helicopter pilots are augmenters or far less frequently, type 2 reducers. As predicted, pilots are also extraverted and much less neurotic than nonpilots. Surprisingly, pilots did not score high in sensation seeking even though augmenters tend to be sensation seekers[4,9] and augmenters, extraverts, and sensation seekers have all been linked with low cortical arousal.[3,4,9,15] The selection process for pilots appears to favor emotionally stable individuals rather than people who seek novel, exciting, or dangerous activities. In addition, although pilots performed reasonably well in the keeping track task, they did not perform as well as the comparison group of scientists and engineers.

This suggests that pilots are not selected for their ability to work under high mental workloads and therefore may experience cognitive overload when keeping track of information on several visual displays. The implication of this for selecting pilots and designing displays is obvious. Human factors research is clearly necessary to determine what information is really necessary and how best to present it.[1,2] Finally, there was some evidence that the more extraverted pilots as well as pilots with fewer errors in the keeping track task performed better in the dual task of acquiring targets and piloting the helicopter simulator.

In conclusion, the EP cortical assessment procedure will likely increase in utility as a selection technique with further research refining the classification of reducers. This study also demonstrates the feasibility and importance of a collaborative basic and applied research program in enhancing progress in both domains toward solving Army problems.

References

1. Remington, R., & Williams, D. (1986). On the selection and evaluation of visual display symbology: Factors influencing search and identification times. Human Factors, 28, 407-420.

2. Dominessy, M. E., Lukas, J. H., Malkin, F. J., Monty, R. A., & Oatman, L. C. A comparison of information display formats for target acquisition. Submitted to Human Factors.

3. Lukas, J. H. (1987). Visual evoked potential augmenting-reducing and personality: The vertex augmenter is a sensation seeker. Personality and Individual Differences, 8, 385-395.

4. Lukas, J. H., & Siegel, J. (1977). Cortical mechanisms that augment or reduce evoked potentials in cats. Science, 198, 73-75.

5. Lukas, J. H., & Siegel, J. (1977). Aversive noise effects on performance and thalamocortical responsiveness in cats. Physiology & Behavior, 19, 555-559.

6. Lukas, J. H., & Mullins, L. F. (1985). Auditory augmenters are sensation seekers and perform better under high workload (Abstract). Psychophysiology, 22, 580-581.

7. Lukas, J. H., & Mullins, L. F. (1988). Evoked potential technique for predicting performance under high mental workloads. Proceedings of the Army Science Conference, 2, 203-215.

8. Buchsbaum, M. S.   (1976).   Self-regulation of stimulus intensity: Augmenting/reducing and the average evoked response.   In G. E. Schwartz & D. Shapiro (Eds.), Consciousness and self-regulation: Advance in research and theory (Vol. 1), (pp. 101-135).

9. Zuckerman, M., Murtaugh, T. M., & Siegel, J.   (1974).   Sensation seeking and cortical augmenting-reducing. Psychophysiology, 11, 535-542.

10. Saxton, P. M., Siegel, J., & Lukas, J. H.   (1987).   Visual evoked potential augmenting/reducing slopes in cats.   2.   Correlations with behavior. Personality and Individual Differences, 8, 511-519.

11. Mullins, L. F., & Lukas, J. H.   (1984).   Auditory augmenters are sensation seekers - if they attend the stimuli (Abstract). Psychophysiology, 21, 589.

12. Eysenck, S. B. G., Eysenck, H. J., & Barrett, P.   (1985).   A revised version of the psychoticism scale. Personality and Individual Differences, 6, 21-29.

13. Zuckerman, M.   (1979).   Sensation seeking: Beyond the optimal level of arousal. Hillsdale, NJ: Lawrence Erlbaum Associates.

14. Lukas, J. H., Siegfried, J. B., & Mullins, L. F.   (1987).   Retinal and cortical augmenting-reducing to flash and pattern reversal stimuli.   In R. Johnson, J. W. Rorrbaugh & R. Parasuraman (Eds.), Current trends in event-related potential research (EEG Suppl. 40), (pp. 220-226). Amsterdam: Elsevier Science Publishers.

15. Eysenck, H. J.   (1967).   The biological bases of personality. Thomas, Springfield.

Measurement of Heat Losses From a
Conduit-Type Heat Distribution System

V.J. Lunardini, Ph.D.
U.S.A. Cold Regions Research and Engineering Laboratory
Hanover, N.H. 03755

It has been estimated that the Department of the Army has approximately 3000 miles of energy-carrying conduits and the Department of Defense has approximately double this. A conservative estimate for the cost to the Army of energy losses from these conduits is $90 million each year. Heat losses from buried conduit energy-distribution systems are not known with confidence, especially in regions of seasonal frost. Values used for design are based on calculation procedures that rely on unverified assumptions. The object of this study was to measure the heat losses from a full-scale buried conduit under controlled laboratory conditions. The data and calculated quantities could then be compared to the predicted heat losses using standard calculation methods.

TEST FACILITY

The Frost Effects Research Facility at CRREL has a test section specifically designed for full-scale utility system testing (Fig. 1). This test section has a trench that is effectively 100 ft long, 8 ft wide, and 8 ft deep, which can be heated or cooled on all four sides. The test facility has two separate 50-ft-long bays side by side so that 80 to 90 ft of conduit can be tested. One of the test bays is fully waterproof so that any earth materials contained in the bay can be wet as desired or a water table can be maintained at any depth. Both of the bays were used in these tests.

A full-scale, standard Class A conduit system, used at many government installations, was purchased from RICWIL Piping Systems, Intergy Inc. The conduit consists of a 3-in. nominal ASTM A-53 GR B Sch 40 steel pipe with thermal insulation, an air gap, and an outer conduit coated with asphalt to retard corrosion, of RICWIL Ferro-Shield 10 ga steel to keep out groundwater. The test loop contained two different-sized conduits joined outside the bulkhead to form a single flow loop. The conduit in the waterproof bay contained 1 in. of mineral wool insulation while the section of the conduit in the second bay had 1.5 in. of calcium silicate. These values were chosen by the manufacturer to keep the maximum heat loss to less than 100 Btu/(hr-ft). Thus, during a test run two different insulation systems could be evaluated simultaneously. The installation of the conduit system followed, as closely as possible, the procedure for an actual field burial.
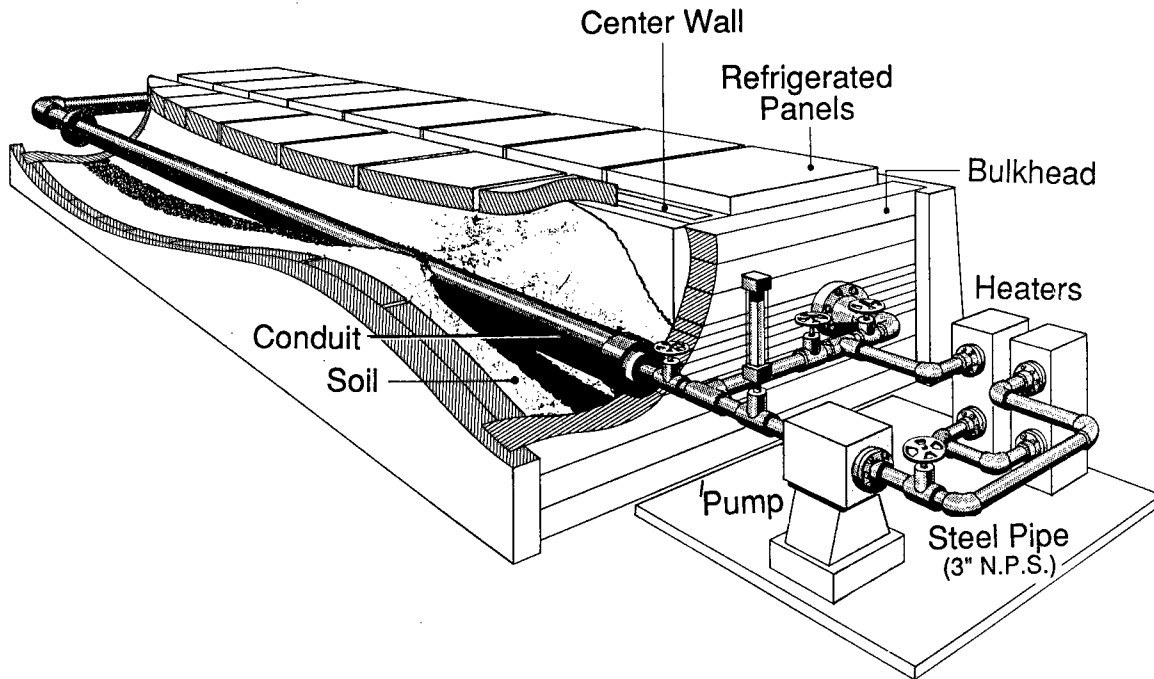
Figure 1. U.S.A. CRREL conduit heat-loss system.

The material used to support and cover the conduit was ordinary sand with some silt and fines. The sand density and moisture were determined by nuclear densometer measurements in situ using a Cambell Pacific Nuclear 501 DR Hydro-probe. The density of the layers varied from 102–140 lb_m/ft^3. The soil moisture in the waterproof section varied from 2.5–21% by weight, while the dry-loop moisture was 2.6–5.4%, varying slowly with time. The kind of sand used was not considered important since the heat losses were measured and field conditions were being approximated.

The flow through the conduit was maintained by a Gould 3196-ST high-temperature 3 HP centrifugal pump capable of flow rates from 5–200 gallons per minute (gpm) and temperatures up to 500°F. The temperature of the system was maintained by two 480 V, 3-phase electrically powered heaters, each rated at 9 kW, which can maintain a maximum water temperature of about 320°F while the pump bearings were maintained below 200°F by a separate water-cooling circuit. The internal pressure of the flowing water was maintained by a nitrogen system such that the water temperatures did not exceed the saturation values.

Each bay has two test sites at which complete heat flux sensor and temperature data were collected. There were from 3 to 5 heat flux sensors located on the surface of the water-carrying pipe at each test site and approximately 16 thermocouple (Type T copper constantan) locations. Four equally spaced thermocouples were epoxied to the surface of the water pipe and another four were inserted just beneath the outer surface of the insulation. The remaining thermocouples at each site recorded the air gap, the outer conduit, and the outside air

temperatures (Fig. 2). During a test run, 13 heat flow sensors and 64 thermocouples were read specifically for the conduit heat loss evaluations at the four test sites. In addition, inlet and outlet water temperatures were measured by nickel resistance thermometers, and about 400 soil temperature thermocouples were monitored. Water flow rates were evaluated with variable area flow meters. All of the sensors were read at the beginning of a test and at the end after a steady state had been reached. A Fluke 2280B Data Recorder and five Fluke 2281A extender chassis were used to record the data, which were stored on magnetic tape and printed on paper for later analysis. Every 4 hours during the entire test duration, 82 values from the heat flow sensors, pipe, insulation, air, and conduit thermocouples, resistance thermometers, and water flow meter were recorded.

x Thermocouples
□ Heat Flow Sensors
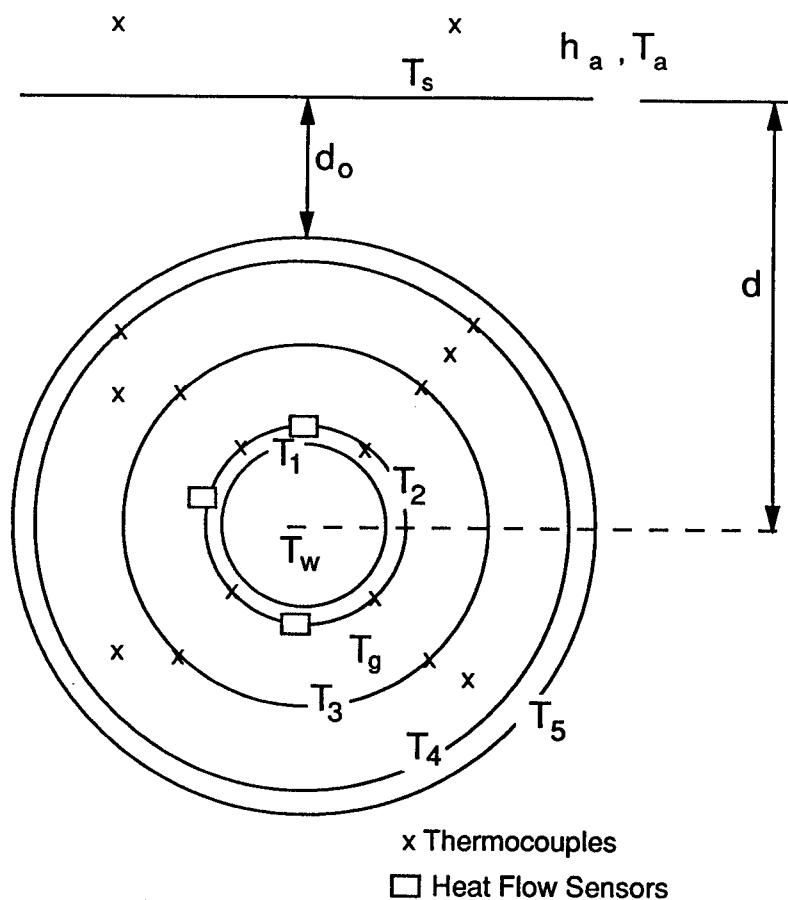
Figure 2. Schematic of buried conduit at a test site.

CALIBRATION AND DATA REDUCTION

Each heat flow sensor (Model C, International Thermal Instrument Co.) was calibrated in situ by comparison with heat-flow sensors of known accuracy that had been used in previous CRREL heat-flow studies (Micro-Foil, RFD Corp.). It was determined that the sensors were sensitive to placement procedures, but the output of the sensors was very stable over long periods of time once they were epoxied in place. It was found that the heat flow sensors varied about 5% simply due to placement on the pipe, i.e., removal of a sensor and reattachment in the same place. The overall accuracy of the heat flow sensors is estimated at ± 8%. The thermocouples were not calibrated but each was checked at the ice point and any that were off by more than 0.1°F were discarded. The pipe thermocouple readings have been found to be very consistent with one another at each test site.

The heat losses from the conduit were determined by four methods: a) heat flux sensors placed on the exterior of the internal pipe, b) pipe and insulation temperature measurements with subsequent calculation, c) water flow and temperature drop measurements (enthalpy method), and d) measurement of energy input to the system and energy losses exclusive of the conduit. The data reduction was carried out using various heat transfer relations and calibration formulae as detailed in Appendix 1.

TEST PROCEDURE

Conduit in Air

After initial testing of the instrumentation, the heat losses from the conduit resting on wooden beams with sand levels at the bottom of the conduit were measured. These tests were used to calibrate the instrumentation as well as to compare the heat-loss measurements using the four methods noted above. The results from the heat flow sensors and the thermocouples (methods a and b) were comparable and quite consistent. Unfortunately the other two methods yielded results that were unacceptable. With the enthalpy method (method c) it was found that, even with flow rates as low as 10 gpm, the water temperatures had to be measured within 0.01°F to give reasonable accuracy. Water temperature measurements with this accuracy were impossible to maintain over long periods of time. The flow rates could be determined to within 1% of full scale but were the least accurate at the lowest flow rates, which yielded the largest temperature differences. Measurements were made to determine the total energy input and the energy losses from each part of the system external to the conduit itself. In principle, the difference between the energy input and the energy losses is the heat loss from the conduit alone. However, the absolute values of the input and losses were far larger than the conduit losses. Thus, even accuracies of 10% in the input and loss values led to unusable estimates for the conduit loss in most cases. It was decided to eliminate these latter two methods in the data analysis.

Conduit in Sand

The next set of tests was carried out with 1 ft of sand covering the top of the conduit. Heat losses for water temperatures up to 320°F were evaluated using the heat flux sensor and the thermocouple data. No water was added to the soil for these tests. The air temperatures about 1 ft above the soil surface were recorded for all of the sites for these and for all of the soil tests. It

LUNARDINI

took 3–4 days to establish equilibrium conditions at each temperature setting. One complete test run for a given set of conditions at several different water temperatures required at least 2–3 weeks. The conduits in each test bay were further covered with sand, in two steps, until about 37 in. of sand lay over the conduit top. The sand was still at its low original water content (< 5%) with no water added, but it was wetter than before due to outside storage. Heat losses were again evaluated for the complete range of water temperatures.

Water was then added to the sand, in the waterproof bay only, until the water table was at desired levels relative to the bottom of the conduit. The water table was determined by measuring the free water level in standpipes built into the test bays and was corroborated by nuclear densometer vertical surveys of the soil moisture and density. Water table levels were maintained up to the middle of the conduit and the heat losses were measured at each water level. The dry bay (no water table), which held the conduit with the calcium silicate, did not change water content during these water-table adjustments and the heat losses measured were thus replicates of the same dry sand test conditions over several months. This allowed the reproducibility of the tests to be evaluated.

RESULTS

Figure 3 shows the heat loss for the mineral-wool-insulated conduit exposed to the ambient air based on the heat flux sensors and the conduction data (heat loss calculated from the thermocouple data [method b]). The results are close to one another and agree with
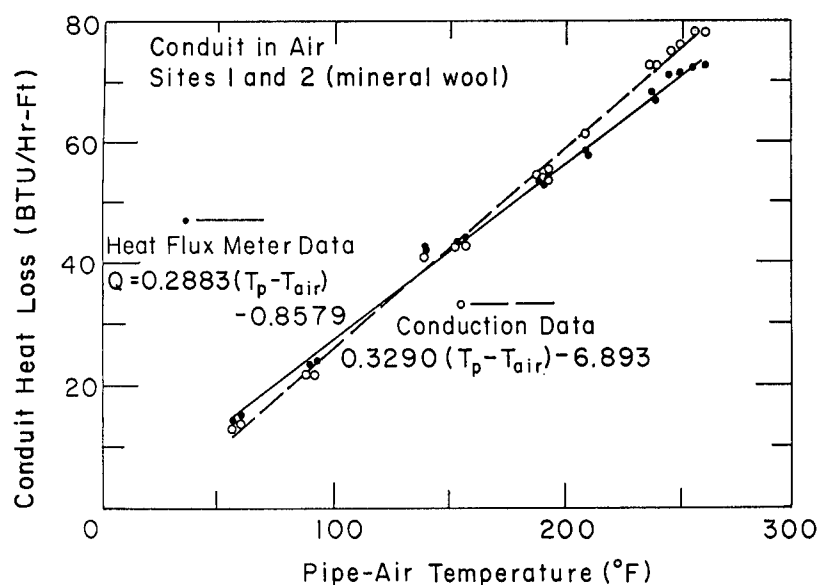
Figure 3. Measured heat loss, conduit in air, mineral wool.

expectations based on the theoretical concepts of Appendix 3. The two methods are within 5.5% at a temperature difference of 250°F. Figures 4 and 5 show the results of all of the tests, including the water table results for Site 1 (mineral wool). As expected, these figures indicate that the heat loss in air is close to the maximum since there is no thermal resistance due to a sand layer; they also indicate that the presence of a water table does not have a major effect on the heat loss. The water table can affect heat loss in two ways: by changing the thermal conductivity of the soil and by introducing convection into the soil. Neither of these phenomena was expected to be significant for the present system and this was justified by the data, although a slight effect of convection may be inferred. The heat loss changes by less than 10% for the range of water table variations noted in Figures 4 and 5. Figures 6 and 7 show very similar results for Site 2 data, also for mineral wool. The measured heat losses at Sites 1 and 2 differed by less than 1%. Figures 8–10 show comparable results for the calcium-silicate-insulated part of the test loop (Sites 3 and 4). Figure 9 shows more clearly the added insulating effect of the soil layers, with the air data prominently separated from the buried conduit data. For Sites 3 and 4 there was no increase of soil thermal conductivity and no soil convective heat transfer effects. Figures 9 and 10 also indicate that the test apparatus and procedures give essentially the same results for the same test conditions run over several months. The heat loss changes by less than 7% from the mean value over a 4-month period. This indicates that the measurements do not drift and should give a true indication of actual physical changes.

A question of some interest for this system is the thermal resistance of the air gap between the insulation and the outer guard pipe. Heat transfer across the air gap is a combination of convection and radiation. The resistance can be evaluated in terms of the effective resistance of
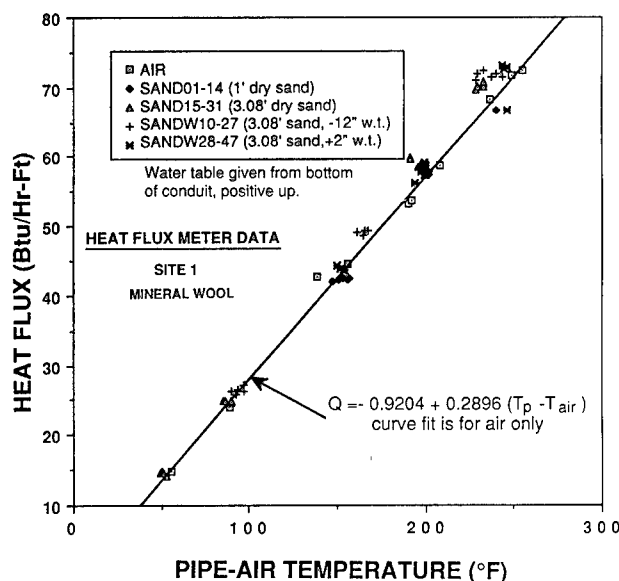
Figure 4. Site 1 heat loss, conduit in sand, mineral wool, heat flux data.
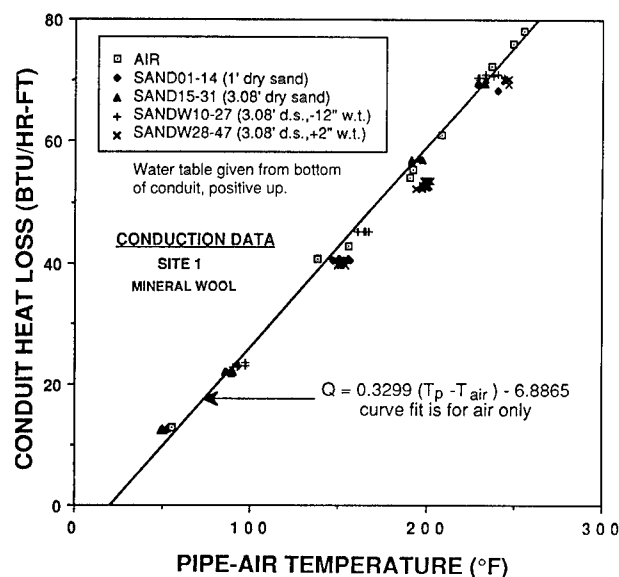
Figure 5. Site 1 heat loss, conduit in sand, mineral wool, thermocouple data.
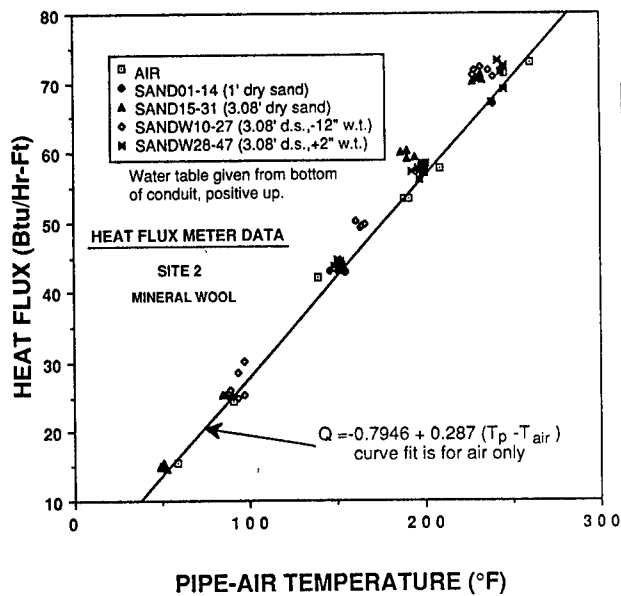
Figure 6. Site 2 heat loss, conduit in sand, mineral wool, heat flux data.
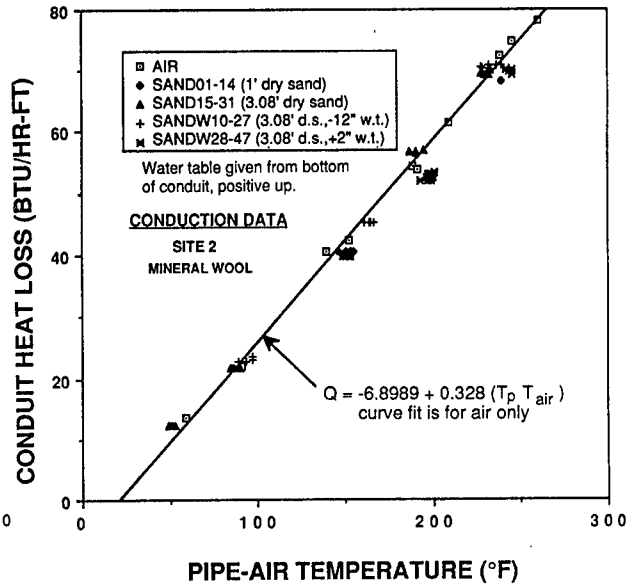
Figure 7. Site 2 heat loss, conduit in sand, mineral wool, thermocouple data.
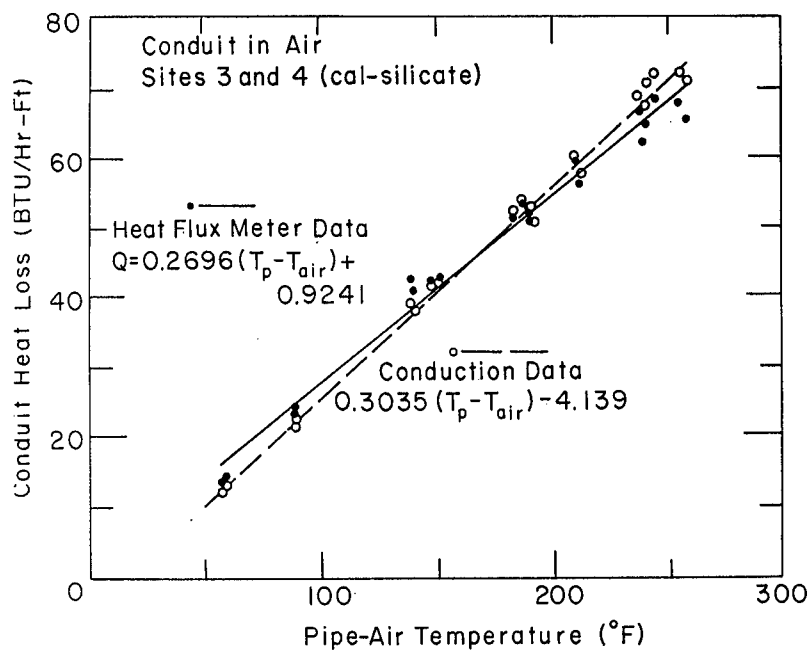
Figure 8. Measured heat loss, conduit in air, calcium-silicate.

441
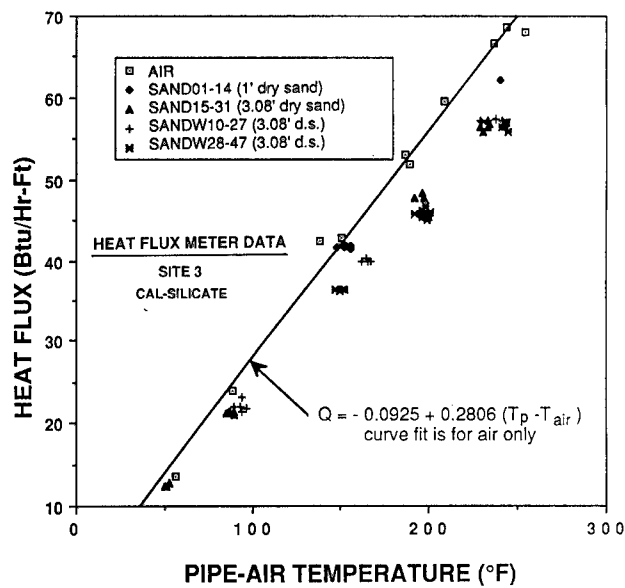
Figure 9. Site 3 heat loss, conduit in sand, calcium-silicate, thermocouple data.
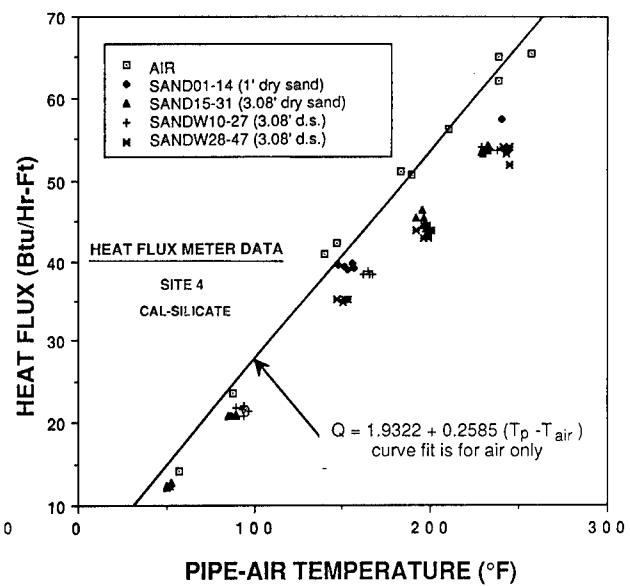
Figure 10. Site 4 heat loss, conduit in sand, calcium-silicate, thermocouple data.
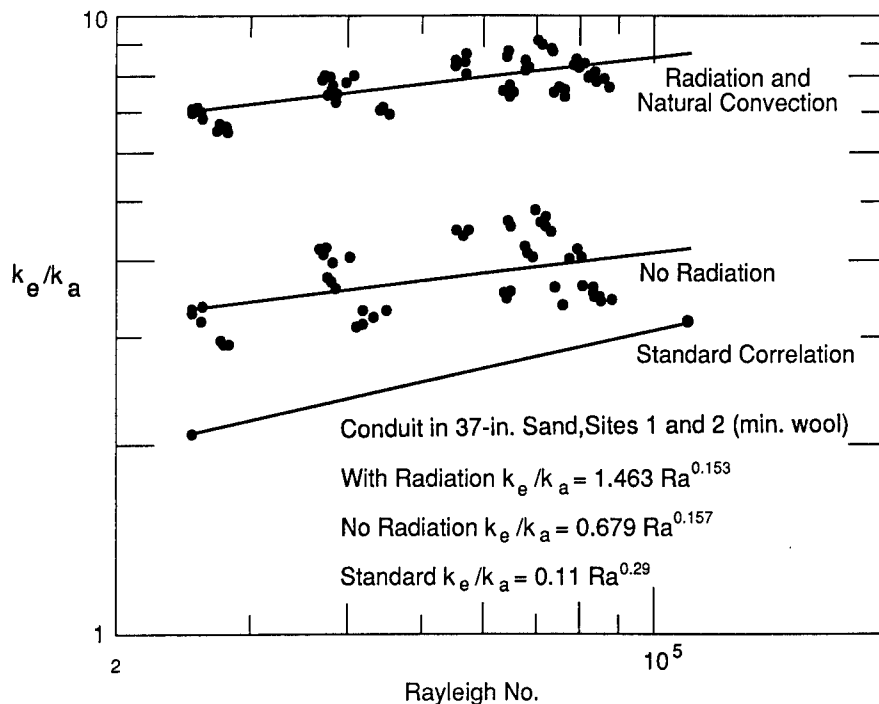
Figure 11. Effective conductive resistance of air gap, all data.

a layer of stagnant air (pure conduction without radiant exchange) within the air space as given by eq 6, Appendix 2. The data for mineral wool are plotted in Figure 11, where the radiation curve is the measured total convective and radiative heat flows across the gap. It can be seen that the data correlation has a slope similar to the standard convective heat transfer correlation for this type of air gap but the constant is considerably higher, indicating that radiation is a significant part of the total heat transfer. The no-radiation curve is an attempt to remove the effect of radiation and recover the convective relation only. In this case the measured values of k_e/k_a approach the correlation of Grober et al. (eq 8).

CONCLUSIONS

Analysis of the data indicates that the heat flux sensors and the pipe thermocouples give similar conduit heat-loss results. Test conditions were maintained over several months without drifting. The enthalpy and energy input methods of measuring conduit heat loss are far less accurate and cannot be used without significantly upgrading the measurement techniques. The measured heat losses are in agreement with the design method values but are in general lower. The effect of a water table up to the middle of the conduit does not appear to cause large increases in the heat losses.

The data from the heat flow sensors and pipe thermocouples have also been compared to calculated heat losses using the standard theory for thermal resistances (see Appendix 3). The standard relations give good results but the thermal resistance of the air gap is underestimated.

ACKNOWLEDGMENT

NOMENCLATURE

A_i, A_c	outer surface area of insulation, inner area of conduit
c_p	specific heat of water Btu/(lb$_m$ – °F)
d_o	depth to top of buried conduit
E	heat flux meter millivolt reading
F	flow rate of water, gallons per minute (gpm)
h_a, h_i	convective coefficients of outside air and water in pipe
k_a	thermal conductivity of air; evaluate at $(T_3 + T_4)/2$
k_c	thermal conductivity of steel conduit (20 Btu/(hr-ft-°F))
k_e	effective thermal conductivity of air gap

k_{g1}	dry soil thermal conductivity, sites 1, 2 (0.5 Btu/(hr-ft-°F))
k_{g2}	dry soil thermal conductivity, sites 3, 4 (0.32 Btu/(hr-ft-°F))
k_i	thermal conductivity of insulation
k_p	thermal conductivity of steel pipe (30 Btu/(hr-ft-°F))
K	heat flux meter constant, Btu/(hr-ft^2-mV)
P_r	Prandtl number, v/α
Q	total heat loss from conduit, Btu/hr.
q	Q/L, Btu/(hr-ft)
r_{pi}, r_{po}	inside and outside radii of pipe
r_{ci}, r_{co}	inner and outer radii of outer conduit
r_i	outer radius of insulation
Ra	Rayleigh number $g\beta\delta^3(T_3 - T_4)/(v\alpha)$
R_T	total thermal resistance of buried conduit
T_{air}	outside air temperature
T_f	$(T_3 + T_4)/2$
T_g	annular gap air temperature
T_p	temperature of pipe, $(T_1 + T_2)/2$
T_s	ground surface temperature
T_w	water temperature
T_{wi}, T_{wo}	water temperature in and out, °F
T_1, T_2	pipe inner and outer temperature
T_3	outer temperature of insulation
T_4, T_5	conduit inner and outer temperature

α	thermal diffusivity of air
β	coefficient of thermal expansion of air
δ	$r_{ci} - r_i$ air gap thickness
$\varepsilon_i, \varepsilon_c$	emmissivity of insulation and conduit
v	dynamic viscosity of air
μ	viscosity of air
ρ	water density, lb/ft^3
σ	Stefan-Boltzman constant, $0.1714 \, \xi \, 10^{-8}$ Btu/(hr-ft^2 – R^4)

LITERATURE CITED

1. Brown, A.I. and S.M. Marco (1958) *Introduction to Heat Transfer,* McGraw-Hill Book Company, New York.
2. Eckert, E.R.G. and R.M. Drake (1959) *Heat and Mass Transfer,* McGraw-Hill Book Company, New York.
3. Data furnished by V. Meyers, CEMRD.
4. Grober, H., S. Erk and U. Grigull (1961) *Fundamentals of Heat Transfer,* McGraw-Hill Book Company, New York.
5. NFGS-15705, p. 15, June 1981.

APPENDIX 1. CALCULATION EQUATIONS

Heat flow sensors. The equation relating the millivolt output of the sensor and the heat loss per foot of pipe is

$$q = 2\pi r_{po} K K_t E. \tag{1}$$

The temperature correction is supplied by the manufacturer (International Thermal Instruments):

$$K_t = 1 - 0.00108 \ (T_p - 75) + 7.3077 \ \xi \ 10^{-7} (T_p - 75)^2 \tag{2}$$

where T is in °F and r is in ft.

Conduction equations. The heat flow through the insulation is given by the standard conduction formula

$$q = 2\pi \ k_i (T_2 - T_3)/\ln(r_i/r_{po}). \tag{3}$$

The thermal conductivity of the insulation material can be expressed as a function of temperature with

$$k = k_o[1 + a(T - 100)] \tag{4}$$

where k is in Btu/(hr-ft-°F), T is in °F, and a is in °F^{-1}. Table 1 lists the values used for the thermal conductivity.

Enthalpy change. The heat loss from the conduit at steady state equals the enthalpy change of the water entering and leaving the conduit:

$$Q = 8.02\rho F c_p (T_{wi} - T_{wo}). \tag{5}$$

Table 1. Values of k_o and a for insulation materials.

Reference	k_o Btu/(hr-ft-°F)		a (1/°F)	
	Cal-silicate	Mineral wool	Cal-silicate	Mineral wool
Brown and Marco[1] (85% mag 13 lb/ft)	0.039	0.028	5.983×10^{-4}	1.9643×10^{-3}
Eckert and Drake[2]	0.034	0.030	5.8824×10^{-4}	1.556×10^{-3}
Epitherm data[3]	0.032	0.023	8.333×10^{-4}	2.174×10^{-3}

APPENDIX 2. HEAT FLOW IN AIR GAP

Effective conduction. The total heat flow in the air gap is made up of convection and radiation. This heat flow can be represented by an effective conduction through a hypothetical, solid material in the annular region. The total heat flow is

$$q = 2\pi \, k_e(T_3 - T_4)/\ln(r_{ci}/r_i). \tag{6}$$

The ratio of the effective thermal conductivity of the gap to that of a layer of stationary air is

$$k_e/k_a = q \, \ln(r_{ci}/r_i)/[2\pi \, k_a(T_3 - T_4)]. \tag{7}$$

The ratio k_e/k_a has been correlated by Grober et al.[4] as

$$k_e/k_a = 0.11 \, Ra^{0.29}. \tag{8}$$

Table 2 gives the relations for the present conduit.

Table 2. Air gap values for conduit (dimensions in ft).

Site	r_{ci}	r_i	δ	r_{pi}	r_{po}	r_{co}	k_e/k_a	Ra
1,2	0.3594	0.2292	0.1302	0.1278	0.1458	0.3698	$0.0716(q/L)/(k_a\nabla T)$	$0.0710\nabla T \, Pr/(T_f v^2)$
3,4	0.4323	0.2708	0.1615	0.1278	0.1458	0.4427	$0.0744(q/L)/(k_a\nabla T)$	$0.1355\nabla T \, Pr/(T_f v^2)$

$\nabla T = (T_3 - T_4).$

Radiation correction: The radiation component of the total heat transfer can be removed with the use of standard heat transfer relations. The radiation across an annular gap is

$$q = \sigma \, F_A \{ 1/[1/\varepsilon_i + (1/\varepsilon_c - 1)A_i/A_c]\}A_i \, (T_3^4 - T_4^4). \tag{9}$$

$$A_i / A_c = \begin{cases} 0.6377 & \text{Site 1, 2} \\ 0.6264 & \text{Site 3, 4} \end{cases} \tag{10}$$

LUNARDINI

The properties of air are taken from Eckert and Drake[2]

$$v = 1.1993 \times 10^{-4} + 6.11 \times 10^{-7} T_f \quad (ft^2/s). \tag{11}$$

$$k_a = 0.0132 + 2.4333 \times 10^{-5} T_f \quad [Btu/(hr\text{-}ft\text{-}°F)]. \tag{12}$$

$$Pr = 0.718 - 0.00012 \, T_f. \tag{13}$$

Thus the purely convective heat transfer across the air gap is

$$q_c = q_t - q_r \tag{14}$$

where q_t is the total measured heat flow. Using eq 7 and 13, a value of k_e/k_a can be calculated, which accounts for only the convective heat transfer. This adjusted correlation can then be compared to the standard relation, given in eq 8.

APPENDIX 3. THEORETICAL HEAT LOSS FROM BURIED CONDUITS

The steady-state heat loss from a conduit is best treated with the use of thermal resistances.

$$q = (T_w - T_{air})/R_T. \tag{15}$$

For the conduit there are seven resistances: 1) convective resistance between the hot water and the pipe (usually small); 2) pipe conduction resistance (usually small); 3) insulation conduction resistance; 4) resistance of the air gap, made up of convection and radiation; 5) conduit conduction resistance (usually small); 6) conductive resistance of the soil layer; and 7) convective resistance at the soil surface. Resistances 6 and 7 will be treated as one resistance. The total resistance is then

$$R_T = \sum_{i=1}^{6} R_i \tag{16}$$

The resistance of each layer is as follows:

$$R_1 = 1/(2\pi \, r_{pi} \, h_i) \tag{17}$$

$$R_2 = \ln(r_{po}/r_{pi})/(2\pi \, k_p) \tag{18}$$

448

$$R_3 = \ln(r_i/r_{po})/(2\pi k_i) \tag{19}$$

The value of k_i is taken at the mean insulation temperature.

$$R_4 = \ln(r_{ci}/r_i)/(2\pi k_e) \tag{20}$$

There are three ways to evaluate the thermal resistance of the air gap, R_4. First, standard correlations for the gap k_e can be used; second, the present experimental data can be used; and third, the current calculation method can be used.

Standard correlation. The values of k_e from Grober et al.,[4] for convective heat flow only, are given by eq 8.

New experimental correlation. The data from the buried conduit experiments have been correlated to give the following equation for the total heat flow across the air gap:

$$k_e/k_a = 1.463 \, Ra^{0.153}. \tag{21}$$

After correcting to eliminate the radiation flux, the convective correlation is

$$k_e/k_a = 0.68 \, Ra^{0.157}. \tag{22}$$

Current calculation method[5]. The current calculation method uses the following relation:

$$R_4 = 1/(6\pi r_i) \tag{23}$$

This equation assumes that the heat flow in the gap is convective with a constant convective coefficient of 3 Btu/(hr-ft^2-°F) based on the surface area of the insulation.

The remaining thermal resistances are:

$$R_5 = \ln(r_{co}/r_{ci})/(2\pi k_c) \tag{24}$$

$$R_6 = \ln(\gamma + (\gamma^2 - 1)^{0.5})/(2\pi k_g) \tag{25}$$

where $\gamma = 1 + (d_o + k_g/h_a)/r_{co}$. If the conduit is not buried

$$R_6 = 1/(2\pi r_{co} h_a). \tag{26}$$

MACDONALD, FRANZ & HEWETSON

Analysis of Ventilatory Patterns During Natural Recovery and
Antibody-Mediated Reversal of Saxitoxin-Induced Apnea
in the Guinea Pig

*Douglas A. Macdonald, Maj., RAVC.
David R. Franz, LTC., VC. and John Hewetson, Dr.
U.S. Army Medical Research Institute of Infectious Diseases,
Fort Detrick, Frederick, MD, 21701-5011, U.S.A.

BACKGROUND

Saxitoxin is a sodium channel-blocking neurotoxin produced by marine dinoflagellates of the Protogonyaulax genus. It is a small, non-protein molecule (372 daltons as the dihydrochloride) which is water soluble, acid stable, and base labile. Saxitoxin is an etiologic agent of paralytic shellfish poisoning, a syndrome caused by the ingestion of saxitoxin-contaminated shellfish, which is characterized by difficult speech; progressive numbness of the lips, face, and extremities; muscular incoordination; respiratory distress; and finally death (ACRES and GRAY, 1978). Mortality in severe cases is 2.6-23.2% and results from respiratory paralysis (SAKAMOTO et al., 1987). The lethal oral dose for a human is reported to be between 0.3 and 1.0 mg and death generally occurs 1-12 hr after ingestion (BOWER et al., 1981). In guinea pigs, the LD_{99} is ~6.0 µg/kg i.v.(FRANZ unpublished data) and, intoxication is characterized by a profound hypotension, neuromuscular paralysis, and respiratory failure. It is generally believed that the cause of respiratory failure in experimental animals is peripheral neuromuscular blockade (KAO et al., , 1967; EVANS, 1969). However, there is debate as to whether the respiratory depression is solely peripheral in origin. Fundamental to this debate is the question of saxitoxin's ability to cross the blood-brain barrier to affect respiratory control. BORISON et al., (1980), conclude from their research that saxitoxin is capable of crossing rapidly; while EVANS (1975) stated that " saxitoxin is unable to pass through the blood-brain barrier." Saxitoxin, a potential biochemical warfare agent (HELMSLEY, 1987), may be used by hostile forces against NATO countries. It is expected that saxitoxin would be delivered as an aerosol and that exposure would occur by inhalation. Saxitoxin is considerably more toxic (LD_{50}~1 µg/kg) when given as an aerosol (CREASIA, 1989). There are no definitive, therapeutic agents available to treat intoxication and the supply of anti-saxitoxin antibody is of animal origin. Therefore, the only effective treatment for intoxication at present is artificial ventilation. Our aim was to understand the respiratory effects of saxitoxin and to evaluate practical approaches to therapy on the battlefield.

MACDONALD, FRANZ & HEWETSON

MATERIALS AND METHODS

All experiments were performed on male, barrier raised, Hartley albino guinea pigs weighing 350-400 g. The guinea pigs were housed in a bio-safety enclosure (Airoclean Engineering, Inc, Edgemont, PA) and received food and water ad libitum. The temperature in the animal room was maintained at 22 ±1°C, with a 12 hr light/dark cycle.

The guinea pigs were anesthetized with urethane (2 g/kg i.p.). Polyethylene cannulae were placed into an external jugular vein (P.E.10) for toxin infusion and a carotid artery (P.E. 60) for arterial blood pressure measurement and blood sampling. A flared P.E. 240 cannula was inserted into the lumen of the trachea after a midcervical transverse tracheostomy and the dead space was adjusted to within normal limits. The guinea pigs were placed in dorsal recumbency into a whole-body flow plethysmograph for measurement of ventilatory flow, respiratory frequency, and tidal volume. Heart rate and blood pressure were measured via a pressure transducer connected to the carotid artery cannula. The physiological variables were calculated electronically (Model-6 Buxco Electronics Inc., Sharon, CT) and confirmed manually from chart recordings. Arterial blood gas tension and pH were measured with an Instrumentation Laboratory System model IL1301 (Lexington, MA). Body temperature was monitored via a thermo-couple probe iserted into the rectum and controlled within normal limits (37-38°C) by placing the guinea pigs on a water-heated pad within the flow plethysmograph. A stock solution of saxitoxin (5 mg/ml in 0.001 N acetic acid; 94% pure by HPLC; provided by Dr. Sherwood Hall, F.D.A., Washington, D.C.) was diluted to 10 µg in 10 µl of phosphate-buffered saline (PBS) and stored at -25°C. Before each experiment, the saxitoxin was further diluted with PBS to a concentration of 10 µg/ml. The dose (ranging from 6.5-15.9 µg/kg) was infused i.v. over 1 min. Apnea typically ensued within 60 sec, at which point, positive pressure ventilation was begun with a rodent ventilator (Harvard Apparatus, South Natick, MA).

Vagotomized Guinea Pigs

A bilateral vagotomy was performed at the time of cannulation. Apnea was induced with saxitoxin (10.2 µg/kg). Tidal volume of the ventilator was adjusted to between 2 and 3 ml to result in a preselected arterial PCO_2 level between 25 and 60 torr, which was maintained within a 5-10 torr range throughout the experiment. Ventilation was interrupted every 10 min to test the guinea pigs ability to breathe independently. The time to the first attempted breath ([1] Fig.1), respiratory frequency during each attempt ([2] Fig.1) (before the onset of hypoxic bradypnea), and the total time on the ventilator ([3] Fig.1) before the guinea pigs were capable of breathing independently were recorded. Two min before we interrupted ventilation, we collected 70-100 µl of arterial blood for the measurement of PO_2, PCO_2, and

pH. If the guinea pig was incapable of sustaining independent breathing, ventilation was reinstated when blood pressure began to fall, apparently in response to profound hypoxia.

Intact (non-vagotomized) Guinea Pigs

The same physiological parameters were measured in guinea pigs with intact vagi. These guinea pigs, used as part of a dose-response study, were challenged i.v. with doses of saxitoxin ranging from 6.5-15.9 µg/kg. Arterial PCO_2 tensions were maintained at pre-toxin levels, typically 30-40 torr, sufficient to drive ventilation. The ventilator was turned off at 20-min intervals to determine the point at which the guinea pigs were capable of breathing independently.

Antibody Therapy

The effect of therapeutically administered antibody was studied in four groups of guinea pigs. A dose of 12.7 µg/kg of saxitoxin was given, as this dose consistently produces apnea lasting for at least 2 hr. The anti-saxitoxin antibody used was burro, polyclonal IgG, purified on a protein G column (Genex Corporation, Gaithersburg, MD). It has an in-vivo neutralizing capacity in mice of 6 mg of antibody to 1 µg of toxin and an affinity for saxitoxin (7×10^{-9} M) (HEWETSON, unpublished data) close to the affinity of saxitoxin for the sodium channel (Kd 1×10^{-9} M) (Krueger et al. 1979). Freeze-dried antibody was dissolved in PBS and warmed to 37°C before i.v. infusion. The antibody was infused over 1 min, beginning 30 sec, 1 hr, or 2 hr after the toxin infusion. A fourth group (anesthetic control) was treated as was the 30-sec group but maintained under anesthesia for 2 hr before the administration of toxin. We selected minute ventilation as being the best indicator of respiratory performance. Measurements of minute ventilation were taken 40, 30, and 20 min before toxin infusion and averaged for each group, to serve as control values. With the ventilator turned off, minute ventilation at 1 min, 3 min, 5 min, and at each succeeding 5 min after antibody administration was contrasted with the pre-toxin control values, to quantify the antibody's ability to reverse the signs of intoxication. Arterial PCO_2 tensions were maintained at pre-toxin levels during the period of artificial ventilation.

Oxygen Therapy

Guinea pigs were studied to determine if increasing the oxygen content of their inspired gases would alter the time of artificial ventilation required before they could breathe independently. Saxitoxin (10.2 µg/kg) was given i.v. to induce apnea and the guinea pigs divided into two groups as follows: 21% group = ventilation with room air from the onset of apnea and throughout the recovery period and 100% group = ventilation with 100% oxygen from the onset of apnea and throughout the recovery period. Oxygen was drawn into the

ventilator from a reservoir bag. Core body temperatures in this study were maintained between 35 and 38°C.

Statistical Analysis

Statistical analysis, as appropriate, was performed by using the SAS Institute IMC, SAS/STAT Users Guide, Release 6.03 edition.

RESULTS

Apnea typically occurred 10-20 sec after completion of the saxitoxin infusion (10.2 µg/kg) and was preceded by hypotension. Not until 20-30 min of artificial ventilation had elapsed did the guinea pigs make any attempt to breathe. Each subsequent 10 min, their breathing became increasingly more effective until they were capable of breathing independently 65 ± 7.83 min after toxin infusion.

MACDONALD, FRANZ & HEWETSON
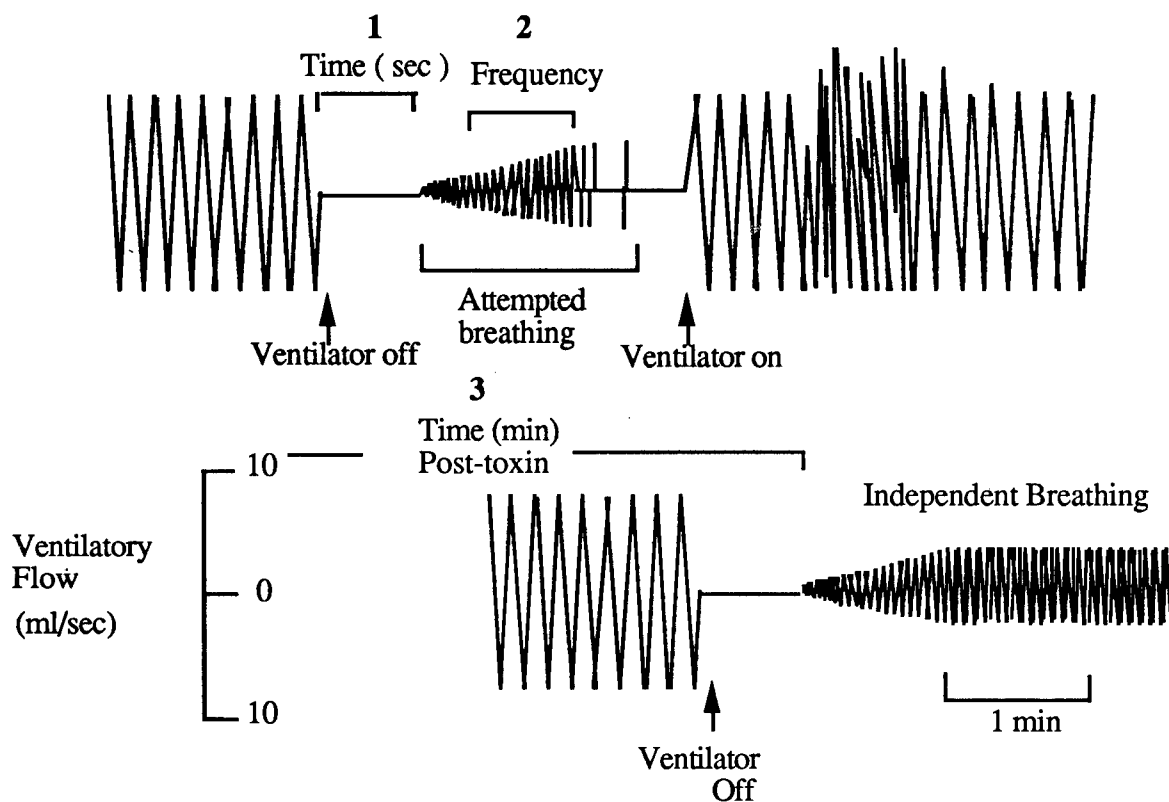
Figure 1. *Stylized representation of a typical ventilatory flow tracing from an anesthetized, saxitoxin-intoxicated, apneic guinea pig supported by artificial ventilation. The upper trace depicts an unsuccessful attempt to sustain independent breathing when the ventilator was turned off. The time to the first breath [1] and the frequency during the attempt [2]were correlated with P_{CO_2}, P_{O_2}, and pH 2 min before the ventilator was turned off and with the total time elapsed after the infusion of toxin. The lower trace demonstrates the point at which the guinea pig was able to breathe independently; the time to independent breathing [3] was correlated to mean P_{CO_2}, P_{O_2}, and pH throughout the period of apnea.*

There was an inverse correlation between arterial P_{CO_2} tensions and the time to the first spontaneous breath. (Fig 2).

Neither respiratory frequency during the periods of attempted breathing (R=0.32), nor the total time of artificial ventilation required before the animals could breathe independently (R=0.01), was altered by arterial P_{CO_2} tension within the range tested. Figure 3 demonstrates that frequency during the periods of spontaneous breathing was characteristic of the individual test animal. Frequency was not related to arterial P_{CO_2} tension.

MACDONALD, FRANZ & HEWETSON
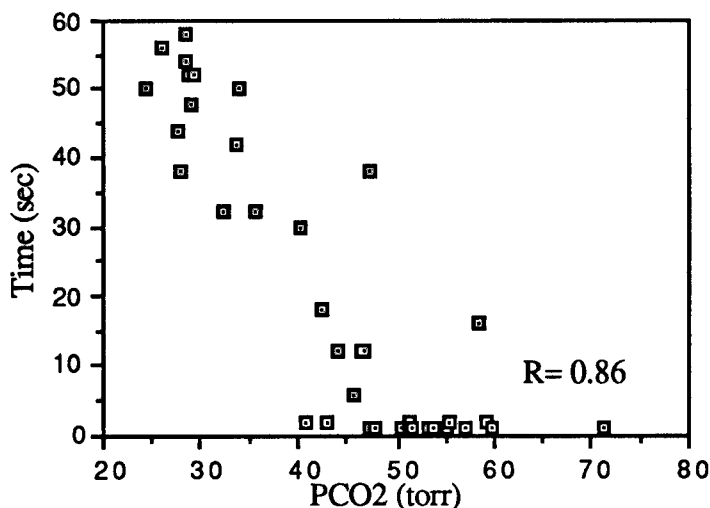
Figure 2. *The time to the first breath compared to arterial P_{CO2} in apneic, saxitoxin-intoxicated guinea pigs. Time was measured from the point that the ventilator was turned off until the first spontaneous attempt to breathe was seen (see [1], Fig. 1).*
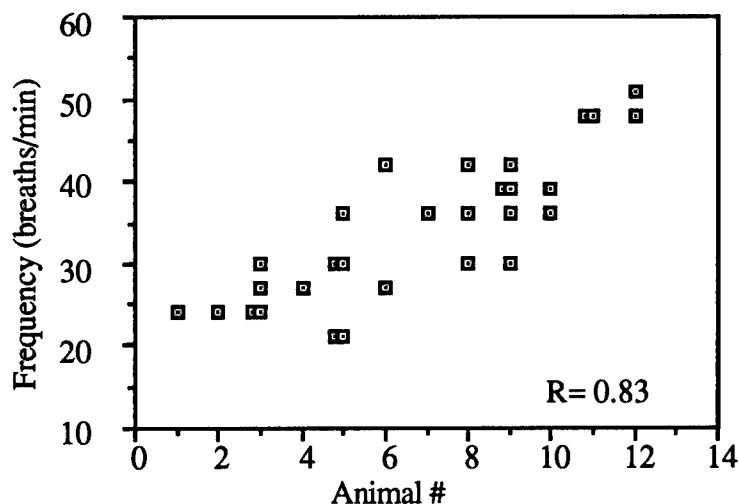
Figure 3. *Respiratory frequency during failed spontaneous attempts (see [2], Fig. 1), grouped by animal number, in guinea pigs rendered apneic by saxitoxin. Individual animal data were ranked artificially from lowest to highest mean frequency.*

Figure 4 shows the respiratory frequency in non-vagotomized guinea pigs during periods of spontaneous breathing when the ventilator was turned off. The trend was to greater within-animal variation at lower doses of toxin. Also, frequency during spontaneous breathing

increased with time after toxin in low-dose animals (6.5-10.2 µg/kg) and decreased in the high-dose (12.7 and 15.9 µg/kg) animals.
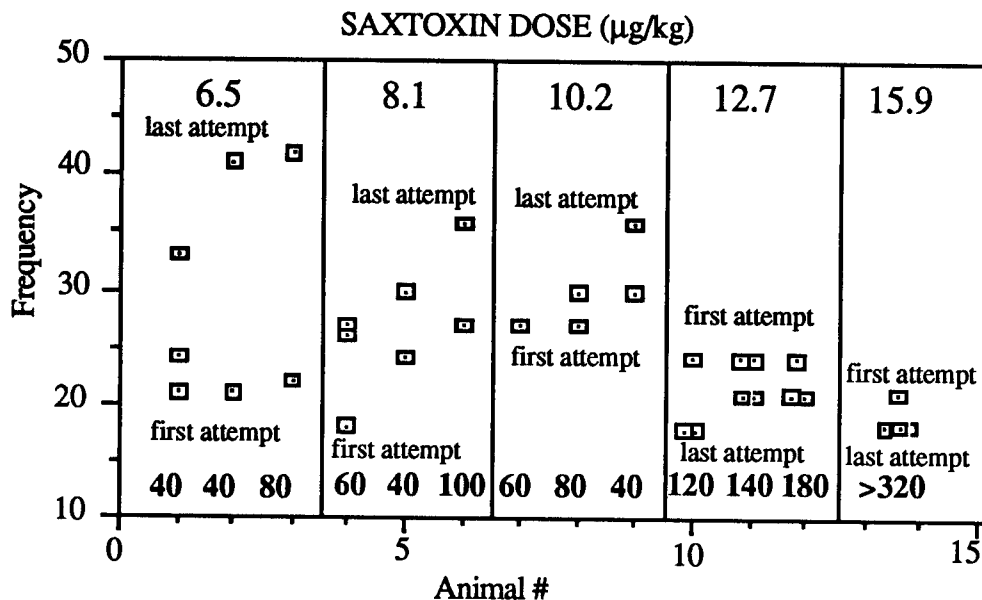
Figure 4. *Respiratory frequency during failed attempts to breathe independently (see [2], Fig. 1), grouped by animal number (x axis) and saxitoxin dose (µg/kg; above the data points), in intact guinea pigs rendered apneic by saxitoxin. Total time (min) of artificial ventilation required is shown above the x axis, in bold, below the data points. Note that respiratory frequencies during the spontaneous breathing increased from first to last attempt in the low-dose (6.5-10.2 µg/kg) animals and decreased in the high-dose (12.7-15.9 µg/kg) animals.*

Antibody Therapy

Antibody therapy alleviated the symptoms of intoxication in all four groups of guinea pigs. Recovery was swiftest in those guinea pigs given antibody 30 sec after toxin, i.e. the control and 30-sec groups. Typically, the ability to breathe independently returned within 1 min of completion of the antibody infusion, and probably occurred before the ventilator was turned off. Due to this limitation of the experimental model we were unable to determine the exact point at which the guinea pigs could breathe independently. One hr after toxin, antibody-mediated reversal of apnea required 3.25 min ± 2.25; and 2 hr after, 5.0 min ± 3.37. There was no statistically significant difference in tidal volume between the groups throughout the experiment, before saxitoxin or after antibody. The minute ventilation of both the control and 30-sec groups did not differ significantly from each other throughout the experiment (p< 0.05) nor did they differ after antibody administration when contrasted to their pre-toxin values (p=0.10). The minute ventilation of the 1 hr group was significantly lower (p<0.05) than the average level of control and 30-sec groups from 1-10 min after antibody administration. Two-hours delay between toxin and antibody treatment resulted in a

significant lowering of minute ventilation ($p<0.05$) from the average level of control and 30-sec for up to 30 min after antibody. Thus, a statistically significant depression of minute ventilation was seen and lasted up to 10 min when the antibody was given 1 hr after toxin and 30 min when given 2 hr after. In contrast, significant depression was not seen in the groups given antibody 30 sec after toxin, i.e., the 30-sec and control groups (Fig. 5). Five minutes before the antibody was given to the 1-hr and 2-hr groups, the ventilator was turned off to ensure the guinea pigs remained apneic. In the 1-hr group, the guinea pigs made no attempt to breathe, while in the 2-hr group, the guinea pigs attempted to breathe, then failed, producing a trace like that illustrated in Fig. 1, [2].
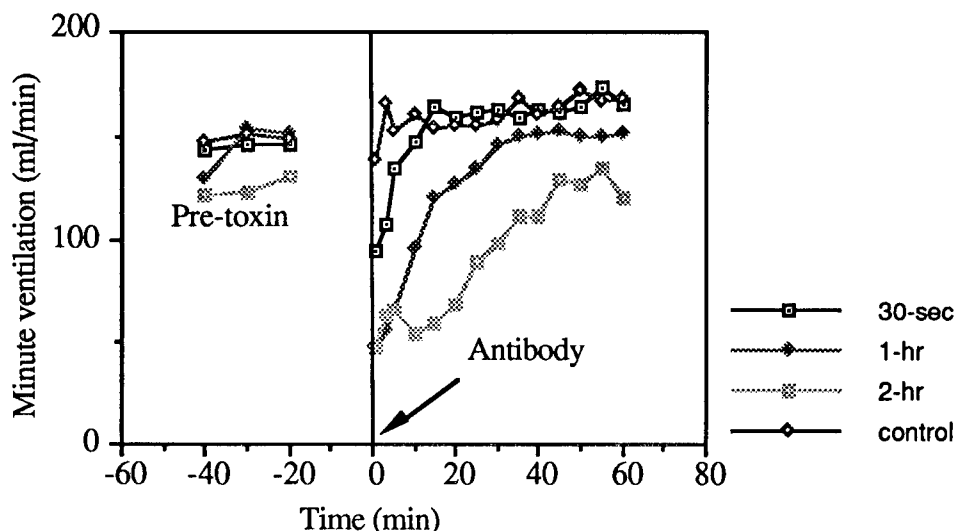
Figure 5. *Minute ventilation before toxin and after antibody. Minute ventilation was significantly depressed ($p<0.05$) when compared with pre-toxin levels for up to 10 min after antibody in the 1-hr group and up to 30 min ($p<0.05$) in the 2-hr group.*

A trend toward depressed minute ventilation was noted immediately after antibody administration in the control and 30-sec groups. The reduction was most marked in the 30-sec group, and lasted between 5 and 10 min; it was slight in the control group and appeared to last for only 1 min. In all the groups the reduction in minute ventilation was caused mainly by a reduction in respiratory frequency. Return of respiratory frequency is illustrated in Fig. 6.
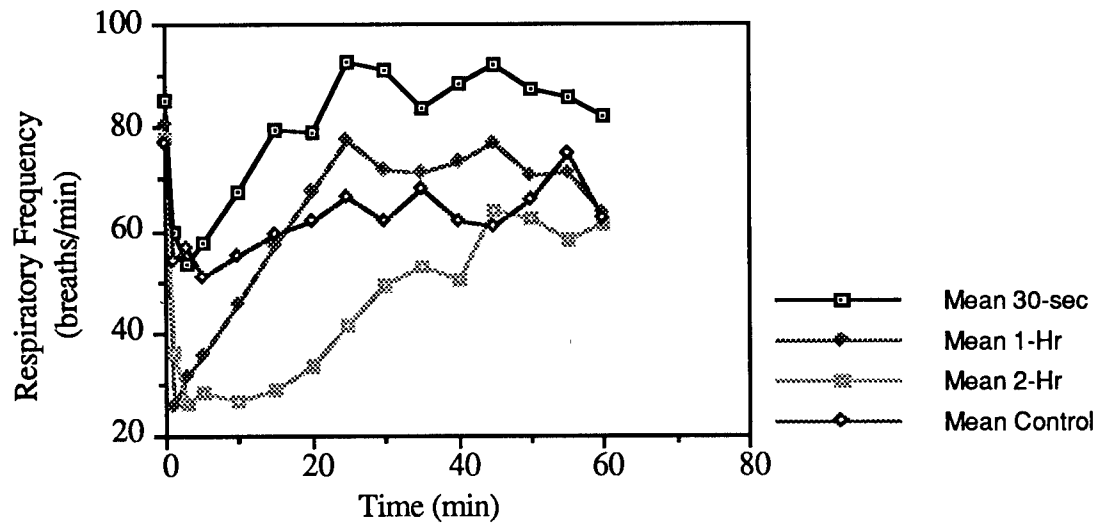
Figure 6. *Respiratory frequency after antibody administration. Time point zero is the average of three time points taken before toxin administration. Note that frequency was depressed in all the groups immediately after antibody was given.*

MACDONALD, FRANZ & HEWETSON

Oxygen Therapy

The substitution of 100% oxygen for room air in the inspired gases of guinea pigs significantly (P<0.05) shortened the duration of artificial ventilation required before they were capable of breathing independently (Fig. 7). There was also a trend to suggest that a lower body temperature prolonged the duration of ventilation required. This temperature effect accounted for much of the variation present in the data.
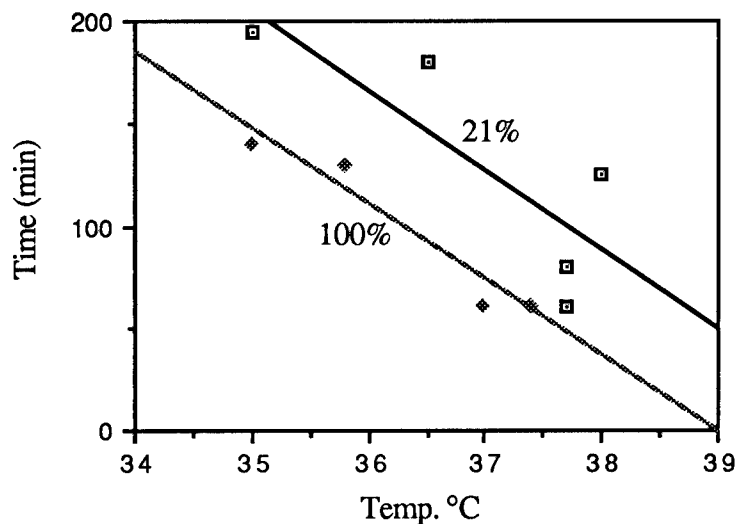
Figure 7. *The duration of artificial ventilation required with room air (21%) and 100% oxygen versus core body temperature. 100% oxygen shortened the duration of ventilation required (p<0.05). A parallel line bioassay was used to establish the relative potency of the two levels of oxygen concentration in the inspired gases and to correct for the variations in core body temperature.*

DISCUSSION

The efficacy of artificial ventilation in the resuscitation of saxitoxin-intoxicated patients has been demonstrated clinically in cases of paralytic shellfish poisoning (ACRES and GRAY, 1978). These anecdotal reports suggest that more than 24 hr of assisted ventilation may be required in humans before recovery ensues, the time probably related to the quantity of toxin ingested, rate of absorption from the gastrointestinal tract, and the metabolism or excretion of

the saxitoxin. We are not aware of any detailed information regarding breathing patterns in patients recovering from saxitoxin-induced apnea.

In the vagotomized guinea pigs, we measured ventilatory flow, a manifestation of the interactions of central and peripheral respiratory control systems and the neuromuscular pump, which produces the pressure gradients driving alveolar ventilation. We did not measure neural activity directly, but we can make inferences regarding the state of function of the neural control systems, based on the peripheral effectors' response, when present.

High arterial PCO_2 tensions were associated with a shorter time to onset of breathing each time the ventilator was turned off. PCO_2 is considered to be the most important stimulus in the control of breathing. Central chemoreceptors, sensitive to PCO_2, indirectly via hydrogen ion concentration, are bathed in cerebrospinal fluid on the ventral surface of the medulla. The aortic and carotid bodies, thought to be of less significance in CO_2 modulation of breathing, are located outside the blood-brain barrier and modulate breathing by sending impulses via the vagus and glossopharyngeal nerves to the respiratory control centers in the central nervous system (CNS). Both sets of chemosensors regulate breathing through central mechanisms. These data from the vagotomized guinea pigs demonstrate that the animals were capable of sensing and responding to arterial PCO_2 concentrations during the time that they were incapable of breathing independently.

Respiratory frequencies measured during the failed attempts to breathe, within a given vagotomized animal, were similar, no matter when measured or what the arterial PCO_2 tension. Respiratory frequency generation is complex, resulting from the interaction of medullary cells associated with inspiration and expiration, and influenced by impulses coming from nuclei in the pons, hypothalamus, and reticular-activating system, as well as afferents in the glossopharyngeal and vagus nerves. The respiratory frequencies recorded during the vagotomized guinea pigs' attempts to breathe were characteristic of the individual animal only and showed no trend to either increase or decrease with time after toxin. Vagotomy in the guinea pig produces a reduction in respiratory frequency. Respiratory frequency increased during recovery in the intact guinea pigs given saxitoxin in doses up to 10.2 µg/kg and decreased with time after toxin in those given 12.7 and 15.9 µg/kg. This disparity, we hypothesize, was caused by the central effects of the toxin becoming more important after longer durations of intoxication. At the low-dose levels, and shorter durations of intoxication, the effects on frequency were mainly peripheral and were associated with blockade of the phrenic and vagal nerves, and respiratory muscles.

Our hypothesis that the central actions of the toxin become more important with increased duration of exposure to the toxin is further supported by the data from the antibody experiments. Respiratory depression, characterized by reduced minute ventilation, increased in the guinea pigs as the duration of exposure to toxin increased; i.e. depression was greater after 2 hr than it was after 1 hr. Yet, after 1 hr and 55 min the guinea pigs were showing signs of recovery, as evidenced by their attempts to breathe, suggesting that a substantial amount of the toxin must have either been excreted or metabolized, leaving less to be neutralized by the antibody. We hypothesize that saxitoxin crossed the blood-brain barrier; after 1 hr enough of the toxin had moved along its concentration gradient from the blood into the CNS to depress respiration for up to 10 min after antibody administration, and after 2 hr a greater quantity of toxin had crossed, depressing respiration for up to 30 min after antibody.

It is known that only picogram quantities of saxitoxin are required to significantly affect respiratory patterns when injected intracerebrally (BORISON et al., 1977).

Our data show that substitution of pure oxygen as the inspired gas shortened the duration of artificial ventilation required before the guinea pigs were capable of breathing independently. This, we hypothesize, was not due to an increase in the rate of excretion or metabolism of the saxitoxin, but to assisted oxygenation of the respiratory muscles at a time when minute ventilation was unable to prevent respiratory muscle fatigue from hypoxia. Initially, animals (intact, 10.2 µg/kg) achieved adequate minute ventilation by increasing the tidal volume while frequency remained depressed. Such a response might result if the peripheral neuromuscular pump was functioning adequately, as witnessed by the normal or increased tidal volume values, but frequency, controlled centrally, was reduced. The frequency could have been altered by direct central action or partial blockade of the vagus, or both. Saxitoxin-infused i.v. blocks transmission in the vagus before it blocks transmission in the phrenics (FRANZ et al.,, 1989). Saxitoxin acts on nerves similarly to local anesthetics in that the last nerves to be blocked and the first to recover are the motor nerves, e.g., the phrenics and intercostals. This explains why the peripheral neuromuscular pump can function adequately while the vagus remains blocked. Though it was not statistically significant, the trend to suggest that a lower core body temperature delays recovery from intoxication may be important (Fig. 7). It emphasizes that body temperatures must be maintained within normal limits to ensure optimal recovery. It may also suggest that if saxitoxin is metabolized, the rate of metabolism may be temperature dependent. As total time of artificial ventilation required is dose dependent (MILLER et al.,, 1989) duration of apnea might be shortened if saxitoxin excretion or metabolism could be enhanced. Thus the maintenance of body temperature appears to be important.

To summarize, these data demonstrate that urethane-anesthetized guinea pigs recovering from saxitoxin intoxication were capable of central frequency generation and responded to arterial PCO_2 stimuli at a time when they were unable to breathe independently. Anti-saxitoxin antibody was capable of alleviating the major peripheral symptoms of intoxication but became less effective as the duration of intoxication increased, suggesting that depression of the central respiratory centers became more important as the duration of intoxication increased. Finally, supplementing inspired gases with oxygen shortened the duration of artificial ventilation required before the guinea pigs were capable of breathing independently. We were unable to discern, with this model, whether central failure had occurred at the time that the guinea pigs became apneic.

In conclusion, these data suggest that saxitoxin, in addition to its production of peripheral neuromuscular paralysis, may cause central respiratory depression. Within a military context, it is clear that the peripheral actions of saxitoxin are those of the greatest concern, as paralysis of the peripheral structures appears to take place before any central effects are evident. Battlefield therapy must be directed, in the absence of a specific antidote, to maintaining the casualty by artificial ventilation until the toxin is either excreted or metabolized. After extended periods of artificial ventilation, however, one must also consider the possibility of saxitoxin's central actions becoming increasingly important.

MACDONALD, FRANZ & HEWETSON

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

ACRES, J. and GRAY, J. (1978) Paralytic shellfish poisoning. Can. Med. Assoc. J. **119**, 1195-1197.

BORISON, H.L. and MCCARTHY, L.E. (1977) Respiratory and circulatory effects of saxitoxin in the cerebrospinal fluid. Br. J. Pharmacol. **61**, 679-689

BORISON, H.L., CULP, W.J., GONSALVES, S.F., and MCCARTHY, L.E. (1980) Central respiratory depression and circulatory depression caused by intravascular saxitoxin. Br J. Pharmacol. **68**, 301-309.

BOWER, D.J., REINHOLD, J.H., MATTHEWS, P.A., and HOWDEN, M.E.H. (1981) Nonprotein neurotoxins. Clin. Toxicol. **18**, 813.

CREASIA, D. A. and NEALLY, M. L. (1988). Acute inhalation toxicity of saxitoxin in the mouse. 7th World Congress on Plant and Microbial Toxins, International Society on Toxinology, Stillwater. OK.

EVANS, M.H. (1969) Cause of death in experimental paralytic shellfish poisoning (PSP). Br. J. Exp. Pathol. **46**, 245-253.

EVANS, M. H. (1975) Saxitoxin and related poisons: their actions on man and other mammals. In Proc. First Internat. Conf. on Toxic Dinoflagellate Blooms. ed. LeCicero. J.R., pp. 337-345. Wakefield: Mass. Sci. Technol. Foundn.

FRANZ, D., MILLER, T., and CHANG, T. (1989) Role of vagal afferents in ventilation and breathing patterns during saxitoxin intoxication. FASEB J. **3**, A1158.

HELMSLEY, J. (1987) The Soviet Biochemical Threat to Nato, Macmillan Press Limited, Rusi Defence Studies pp 10-11.

MACDONALD, FRANZ & HEWETSON

KAO, C.Y., SUZUKI, T., KLEINHAUS, A.L., and SIEGMAN, M.J. (1967) Vasomotor and respiratory depressant actions of tetrodotoxin and saxitoxin. Arch. Int. Pharmacodyn. Ther. **165**, 438-450.

KRUEGER, B.K., RATZLAFF, R. W., STRICHARTZ, G. R., and BLAUSTEIN, M. P. (1979) Saxitoxin binding to synaptosomes, membranes, and solubilized binding sites from rat brain. J. Membr. Biol. **50**, 287-310.

MILLER, T. and FRANZ, D. (1989) Effect of mechanical ventilation on saxitoxin intoxication in guinea pigs. FASEB J. **3**, A688.

SAKAMOTO, Y., LOCKEY, R.F., and KRZANOWSKI, J.J., Jr. (1987) Shellfish and fish poisoning related to the toxic dinoflagellates. South. Med. J. **80**, 866-872.

Deformation Behavior and Its Relationship to the Penetration
Performance of High-Density KE Penetrator Materials

Lee S. Magness and Timothy G. Farrand
Ballistics Research Laboratory
Aberdeen Proving Ground, MD. 21005-5066

I. INTRODUCTION

Tungsten heavy (WH) alloys and depleted uranium (DU) alloys are the primary high-density materials used today in long rod kinetic energy (KE) ammunition. In one-to-one comparisons against the spectrum of heavy armor targets, the DU alloys have consistently demonstrated superior terminal ballistic performance. A model scale ballistic test program was conducted at the Ballistics Research Laboratory (BRL), to investigate the variety of fundamental differences in material characteristics of the two alloy systems and their possible role in uranium's superior ballistic performance.

II. BACKGROUND

2.1 The Metallurgies and Mechanical Properties of Uranium and Tungsten Heavy Metal Alloys.

WH alloys and DU alloys are two very different kinds of materials. The uranium alloys are relatively homogeneous, cast materials, that are subsequently worked and/or heat-treated to develop the desired mechanical properties. Pure uranium has an orthorhombic crystal structure at room temperature (alpha phase), but undergoes two phase transitions before reaching its melting temperature of 1132° C. The standard DU alloy for U.S. KE penetrator applications is the binary U-3/4Ti alloy. The titanium is put into solution in the high-temperature body-centered cubic (gamma) phase, which is then rapidly quenched, producing the lenticular martensite microstructure shown in Figure 1a. An aging heat-treatment strengthens the alloy while reducing its ductility, through the precipitation of an equilibrium U2Ti phase. The standard heat-treatment for large caliber KE ammunition (M833 Spec.) is 16 hours at 360° C or 4 hours at 385° C.

Tungsten is a refractory metal, having a body-centered cubic crystal structure with no phase transformations between room temperature and its melting point of 3410°

C. Due to tungsten's poor ductility and toughness, and the high temperatures required to fabricate tungsten parts, the WH alloys used for K.E. penetrator applications are tungsten particle - metal matrix composites, usually produced by liquid-phase sintering. A porous compact of fine powders of tungsten and of the matrix elements (usually only nickel and iron, but additions of copper, cobalt, etc., may also be included) is placed in a furnace and held at temperature a sufficient time for the tungsten to dissolve into the liquified matrix. During cooling, much of the tungsten re-precipitates out of solution, resulting in a micro-structure of fine, nearly pure tungsten particles (30 to 60 microns in diameter) embedded in a matrix alloy of primarily nickel, iron, and a large percentage of retained tungsten (Figure 1b).

A   B
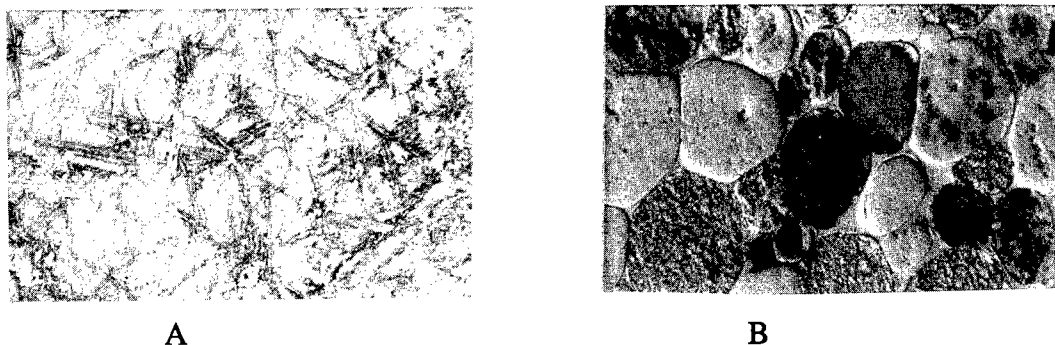
Figure 1.   Typical Microstructures of Uranium - 3/4 Titanium Alloy (A) and Tungsten Heavy Alloy (B).

The static mechanical properties, including Rockwell "C" hardness, yield strength at 0.2% offset (0.2% YS), ultimate tensile strength (UTS), and the tensile elongation (Elong), of several representative uranium and WH alloys are listed in Table 1. Conventional WH alloys lack the precipitation-hardening mechanism of the U-Ti alloys. Cold-working, usually swaging, develops the various combinations of strength, hardness, and ductility in these alloys. The ductilities and toughnesses of these alloys generally decrease with increasing tungsten content, due to the decrease in the volume of ductile matrix surrounding the particles, and the increasing frequency of the contacts between the tungsten particles. These contacts are sources of microcracks and low energy fracture paths in the composite.

The mechanical properties of several heat-treatments of U-3/4Ti and several ternary uranium alloys are also listed in Table 1. The static properties of the best 90-93% W content WH alloys can approach those of the U-3/4Ti alloys, but the densities of these WH alloys are significantly lower. At a tungsten content of 97% (by weight), the density of the WH alloy now equals that of the U-3/4Ti alloy (18.6 g/cc) but the tensile ductility

TABLE 1. <u>Mechanical Properties of Uranium and Tungsten Heavy Alloys.</u>

Material	Density (g/cc)	Hardn. (Rc)	0.2% YS (MPa)	UTS (MPa)	Elong. (%)	Lim. Vel. (m/s)
97W (Ni,Fe,Co) swg 20%	18.6	44	----	1208	4	1318
96W (Ni,Fe) swg 12%	18.3	40	1000	1037	7	1313
93W (Ni,Fe,Co)swg 25%,aged	17.7	47	----	1455	6	1326
93W (Ni,Fe) swg 18%	17.6	41	1116	1197	13	1338
90W (Ni,Fe) swg 20%	17.2	41	----	----	--	1350
U-.1Be-.7Ti	18.6	41	841	1379	11	1228
U-3/4Ti M833 spec. aging	18.6	40	848	1489	24	1221
U-3/4Ti >aging	18.6	47	1100	1580	12	1203
U-3/4Ti >>aging	18.6	49	1089	1689	6	1193
U-.9Hf-.5Ti	18.4	--	896	1620	20	1221
U-1Mo-.5Ti	18.4	49	1396	1821	6	1177
U-3.5Zr-.7Ti	17.3	54	1764	2051	1	1216
U-.5Be-.7Ti	17.2	(40)	(848)	(1489)	(20)	1228

and toughness, (as measured in unnotched Charpy impact tests), fall far short of those of uranium alloys.

The <u>elastic</u> moduli of the WH alloys are 2 to 3 times higher than those of the DU alloys, however, the uranium alloys work-harden more rapidly than the WH alloys with <u>plastic</u> strain. In dynamic testing, the WH alloys have been found to be more strain-rate sensitive than the DU alloys[1,2], exhibiting greater increases in both yield and ultimate strengths and greater decreases in tensile ductility at elevated strain rates.

2.2 <u>The Penetration Performance of Uranium and Tungsten Heavy Alloy Projectiles.</u>

The process of penetration of a target by an eroding projectile is to the first order a hydrodynamic process. At typical ordnance velocities, the inertial forces generated by the acceleration and displacement of the target material and the deceleration and erosion of the penetrator greatly exceed the strengths of both projectile and armor materials. The penetrator is consumed as it burrows into the armor, "back-extruding" from the penetrator-target interface while the interface moves forward into the target (see Figure 2). The influence of the penetrator's material characteristics (its density, mechanical properties, etc.) on its ballistic performance, have been modelled analytically, by engineering models such as the 1-dimensional eroding rod model[3,4], and computationally, using 2- and 3-dimensional finite element hydrocodes[5]. Many aspects of the penetration process remain difficult to

model however, and many details of the physical and material processes, under the conditions of high pressures (5-50GPa) and high rates of deformation (10^4 - 10^6 s^{-1}), are not well understood.
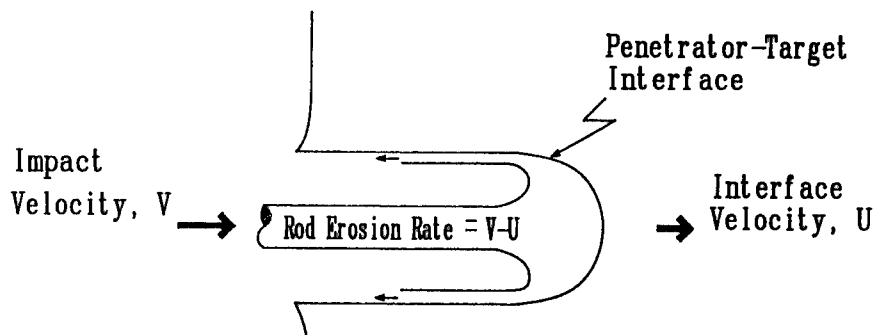
Figure 2. Idealized Hydrodynamic Penetration by a Long Rod.

The influence of material parameters have also been studied experimentally, in model scale and small caliber ballistic tests[6,7,8]. The ballistic database generated at the BRL is probably the most extensive of this kind. Many of the ballistic performance trends seen in these model-scale tests are also observed in full-scale tests, although due to the higher costs of large-caliber testing, the ballistic data is less extensive.

In Figure 3, the penetration depths into rolled homogeneous armor (RHA) steel as a function of impact velocity, of model scale DU and WH alloy long rod penetrators,
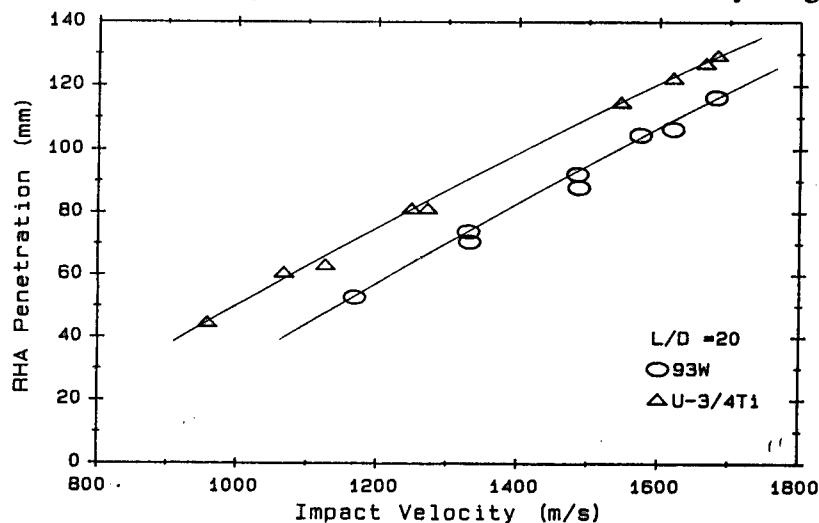
Figure 3. Comparison of the Penetration Performance of WH and U-3/4Ti Alloy Projectiles into Rolled Homogeneous Armor - fixed projectile mass (65 g.), rod length-to-diameter ratios of 20.

are compared. The equal mass, equivalent geometry rods were fabricated using a good quality 93% W content, WH alloy (density of 17.6 g/cc) and the U.S.-standard, M833-specification U-3/4Ti alloy (density of 18.6 g/cc). The delta in the penetration capabilities apparent in Figure 3 narrows only slightly when equal density, 97%W penetrators are tested. Similar differences between DU and WH alloy penetration performance have been noted in numerous BRL and DoD reports[6,7,9].

Another method of quantifying the penetration performance of different alloys is by determining limit velocities, the velocity at which a given projectile just perforates a finite thickness of armor. The last column in Table 1 lists the limit velocities, accurate to within 5 to 10 m/s, of model scale, 65 gram mass long rod (overall length-to-diameter (L/D) ratio of 15) penetrators against a 3 inch thick RHA plate, for this sampling of DU and WH alloys. The DU alloys defeat the RHA plate at velocities approximately 100 m/s or more below those required by the equal density (and geometry) WH alloys. The ballistic performance of the WH alloys do improve slightly with increasing alloy density, due to the greater mass-to-presented area of the rod. Large changes in the nominal mechanical properties of alloys of a given tungsten content, however have shown little effect on the limit velocity[7]. By contrast, in these and in prior tests[8], increasing the mechanical strength and hardness of a DU alloy does appear to significantly improve its ballistic performance.

Perhaps because there are so many fundamental differences in the metallurgies and mechanical properties of WH and DU alloys, there have been several different explanations offered for the superior penetration performance of the uranium alloys. The poorer mechanical properties of the higher density WH alloys or alternately, the lower densities of the 90 - 93% W content alloys with good mechanical properties, have often been blamed. Model scale ballistic data, such as above, however, suggest that neither the density difference nor the differences in engineering properties are the root cause. High-energy x-ray radiographic images of the penetrator-target interface[9], taken during the penetration process, have shown that less of a flare develops on the DU projectile. It was suggested that lower work-hardening with plastic strain in DU alloys might be a contributor to its superior performance. This observation, however, would be a reversal of the work-hardening rates observed in instrumented tests at moderate strain rates[1,2]. Still others have suggested that melted uranium or an even lower melting point U-Fe intermetallic compound, may somehow serve as a lubricant at the penetrator-target interface and thus allow the uranium projectile defeat the target more easily[10]. None of the above models gives a fully satisfactory explanation of the performance gap between DU and WH alloy projectiles or the different ballistic responses of each alloy system to changes in mechanical properties.

3.0 Deformation Behavior and Its Possible Relationship to the Ballistic Performance of DU and WH Alloys.

DU and WH alloy penetrators exhibit a fundamental difference in their deformation behavior during penetration. The residual WH alloy penetrators which are observed exiting the rear face of a target (Figure 4a) or are found embedded in the armor (Figure 4b), usually retain a large mushroomed head which developed during the erosion and back-extrusion of the penetrator rod (Figure 2). It is clear from examinations of the mushroomed head and the discarded chips which line the penetration tunnel, Figure 4c, that a great deal of bulk plastic deformation of the WH alloy occurred before the deformation began to localize and the back-extruded material was finally discarded.

The residual U-3/4Ti alloy penetrators, on the other hand, observed exiting the target (Figure 5a) or found embedded in the armor (Figure 5b), have a chiseled-nose appearance. An examination of the microstructure of the back-extruded material (Figure 5c) shows it to be made up of discrete chips with little plastic deformation within each chip. Most of the deformation is localized, in adiabatic shear bands between the chips.

To explore the possible role of deformation behavior in the penetration process, a series of ballistic tests were conducted, evaluating a more diverse selection of WH and DU alloys. Many of the materials evaluated also served to assess other explanations for the DU vs. WH alloy performance gap. The features of the penetration tunnels into RHA were noted. Many of the residual penetrators found embedded in the armor were metallographically examined. Table 2 lists the mechanical properties and the ballistic limit velocities of model scale L/D = 10 rods against the 3.0 inch RHA target, with additional notes about each material's characteristics and ballistic behavior.

Although the ballistic performances of the WH alloys again improved with increasing alloy density, one notes an approximately 100 m/s performance gap between the standard, orthorhombic crystal class U-3/4Ti alloy and the equi-density WH alloy (97.5% WNiFe). The ballistic performances of the conventional nickel-base matrix WH alloys again appear to be virtually unaffected by alloy strength. All of the penetration tunnels displaced by the U-3/4Ti alloy and the conventional WNiFe alloys were roughened in appearance, due to the irregular discard of eroded rod material as the penetration proceeded. In one WH alloy, however, the nickel-base alloy in the matrix was replaced by tantalum[11], a dense but ductile refractory metal with a melting point of nearly 3400° C. The deformation behavior of the W-Ta composite penetrator, as illustrated in Figure 6a, was quite ductile, almost without localizations of the plastic flow. The discard of the back-extruded material occurred along the periphery of the mushroomed head, which left a remarkably smooth tunnel in its wake. It is interesting to note that a similar ductile flow behavior was also observed for pure tungsten penetrators, in this study and by other
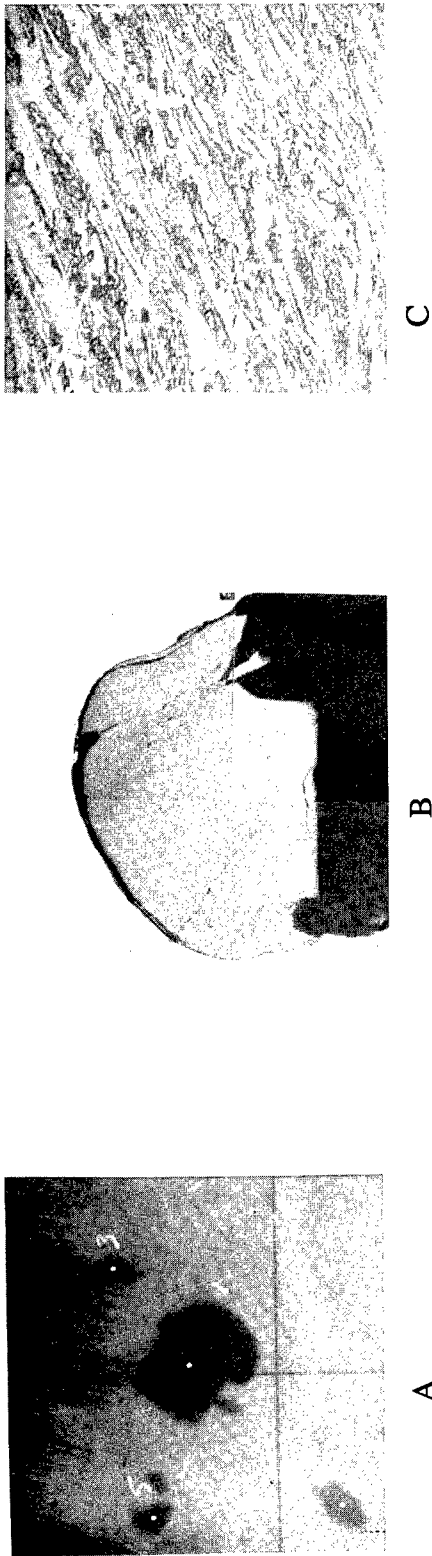
Figure 4. Residual WH Alloy Penetrators: Exiting RHA Target (A), Embedded in RHA Target (B), and Plastic Deformation within Discarded Chips (C).
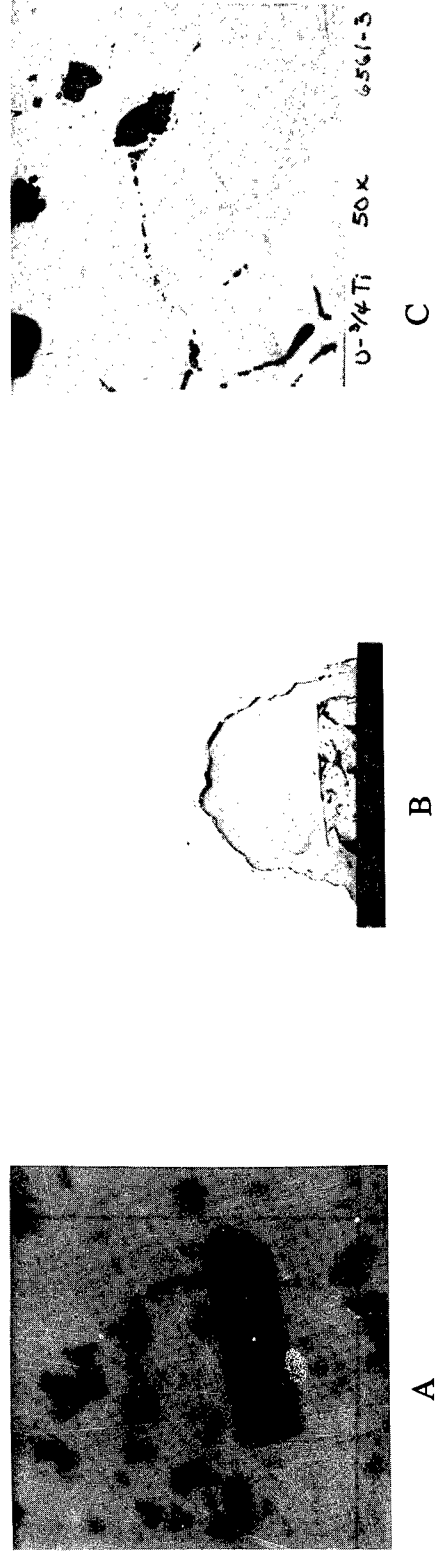
Figure 5. Residual U-3/4 Ti Alloy Penetrators: Exiting RHA Target (A), Embedded in RHA Target (B), and Eroded Rod Material Divided by Adiabatic Shear Bands (C).

TABLE 2. Uranium and Tungsten Alloy Diagnostic Tests

Material	Density (g/cc)	Modulus (GPa)	0.2% Y.S. (MPa)	UTS (MPa)	Comp.S. (MPa)	Lim. Vel. (m/s)	Comment
DU-3/4Ti	18.6	119	800	1440	980	1322	Rough Tun.
(M-833 Spec., alpha-class uranium)							
97.5 WHA	18.6	---	----	1080	----	1430	Rough Tun.
(97% W content WH alloy, Ni-Fe-Co-Cu matrix, swaged 14% RA)							
80W-20Ta	18.6	---	NA	NA	----	1455	Smooth Tun.
(Ta-coated tungsten powders consolidated by hot isostatic pressing)							
96 WHA	18.3	350	----	1103	1240	1453	Rough Tun.
(96% W content WH alloy, Ni-Fe matrix, swaged 12% RA)							
93 WHA	17.6	---	1103	1190	1313	1461	Rough Tun.
(93% W content WH alloy, Ni-Fe matrix, swaged 18% RA)							
91 WHA(CW)	17.3	---	924	1227	----	1470	Rough Tun.
(91% W content WH alloy, compressively cold-worked by upsetting)							
DU-6Nb	17.3	62	275	900	----	1475	Smooth Tun.
(6% niobium content, monoclinic gamma-zero phase uranium alloy)							
90.7 WHA	17.2	386	1180	1258	----	1490	Rough Tun.
(90.7% W content WH alloy, Ni-Fe matrix, swaged 18% RA & strain-aged)							
90 WHA	17.1	362	690	917	----	1495	Rough Tun.
(90% W content WH alloy, Ni-Fe matrix, unworked)							
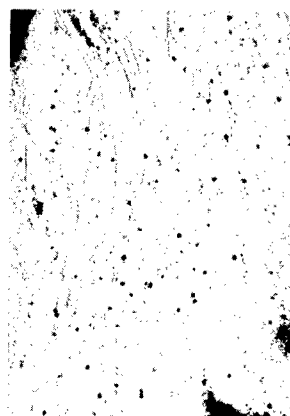
A   B

Figure 6. Microstructures of Residual Penetrators, W-Ta Alloy (A) and U-6Nb Alloy (B).

investigators[12,13]. The limit velocity for the W-Ta material was 25 to 30 m/s higher than that of an equi-density 97% W content, nickel-base matrix composite.

An alloy of uranium and 6% niobium (U-6Nb) was also included. This metastable uranium alloy has a monoclinic structure, is quite soft, and its elastic modulus is even lower than that of the orthorhombic DU-3/4Ti alloy. Its melting point is approximately the same as the DU-3/4Ti, and it can be expected to melt and/or form the same "lubricating" uranium-iron intermetallic compounds at the penetrator-target interface. Unlike DU-3/4Ti, it left a smooth-walled penetration tunnel. No localized shear bands were found in the microstructure of the back-extruded tube (Figure 6b). The uniform microstructure of fine, equi-axed grains suggested that all of the penetrator material underwent bulk plastic deformation and was then re-crystallized. Most significantly, this uranium alloy which did not exhibit adiabatic shear failures, also did not perform as well as the uranium alloys which failed by adiabatic shear. Its limit velocity was very close to those for equi-density WH alloys.

The erosion of the uranium alloy penetrators by adiabatic shear failure has been noted in the past(12,14). However, the implications of the differing deformation behaviors of DU and WH alloys to their ballistic performance have never been explored in any detail. Indeed, because adiabatic shear bands result in an earlier failure of the penetrator material, they have generally been viewed as a detriment in penetration[12]. This study proposes an opposite viewpoint, that the adiabatic shear behavior of uranium aids its performance by minimizing the size of the mushroomed head on the eroding penetrator. The uranium projectile would therefore penetrate more efficiently, making a smaller diameter penetration tunnel, but achieving a greater penetration depth.

To observe the differences in hole profiles postulated above, a special series of tests compared the performance of identical density and dimension U-3/4Ti and 97%W content WH alloy penetrators. To exaggerate the dimensions of penetration tunnels displaced by the small model-scale rods, the tests were conducted against plates of soft mild steel, as well as RHA steel. The profiles of the penetration tunnels (Figure 7), determined by long exposure radiography, verify that at a given impact velocity, the DU rod displaces a narrower, but deeper, penetration tunnel than the WH alloy rod. The differences are especially noticeable at the lower impact velocities. Differences in the deformation behaviors of the penetrator materials may have less of an influence on the flow of penetrator and target materials at the higher impact velocities, due to the dominance of inertial forces. Later in the penetration process, when what remains of the penetrator has slowed down somewhat, a difference in the hole diameters again becomes discernable in the radiographic images. The convergence of the penetration capabilities of the DU and WH alloy rods occurs in both the mild steel and RHA data sets in Figure 7, reflecting the trend seen for the hole profiles. It is interesting to note, however, that the performance of higher

length-to-diameter ratio rods (Figure 3), do not converge very quickly over the span of ordnance velocities.
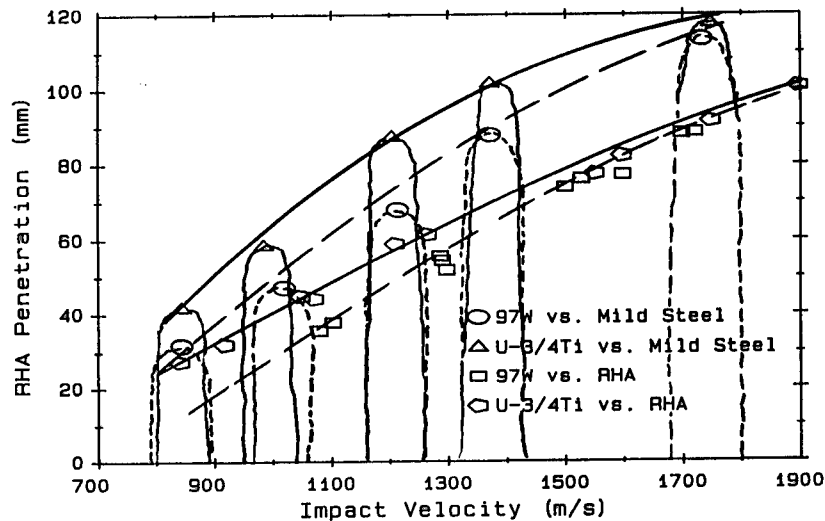
Figure 7.   Comparison of the Penetration Tunnel Profiles of Equal Density, Equal Geometry Rods Into Mild Steel Targets.

4.0   An Adiabatic Shear Susceptibility Model for Ballistic Performance.

Adiabatic shear bands are localizations or instabilities in plastic deformation which develop due to a material's thermal-softening behavior. Approximately 90% of the work put into a material as plastic deformation is converted to heat. Under the high rates of deformation at the eroding end of a projectile during armor penetration, there is very little time for the transport of the generated heat and a rapid rise in temperature results. In materials which soften, due to the rising temperatures, more quickly than they strengthen from strain-hardening and strain-rate hardening mechanisms, the load-bearing capacity of the material will decrease[15,16]. An instability then develops, since any perturbation of the strain field, or region of locally higher plastic strain, will quickly become softer than the material around it, and all further deformation will tend to localize in this weaker region.

The penetration of armor by a projectile is in one sense an ultra-high rate compression test of the penetrator material. Compressive instabilities cannot develop since the local increase in the crossectional area of the rod under compression stabilizes the deformation. Shear localizations can develop, however, on planes angled with respect to the penetrator-target interface, as shown in Figure 8b. This is responsible for both the appearance of the chisel-nose on the residual uranium penetrator and the discard of the back-extruded material as discrete shear chips (Figures 5a-c). The adiabatic shear

localizations develop in the projectile in the same manner as the thermal crosses which appear in steels in high-speed bolt-heading operations[15].
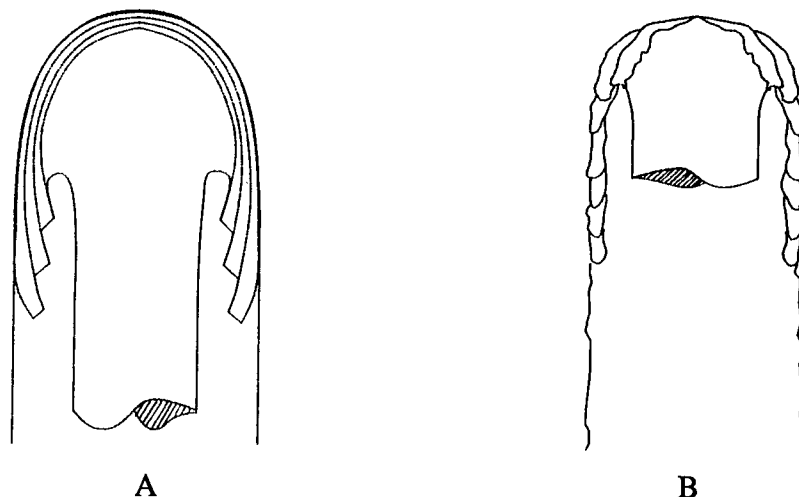
Figure 8. Representation of the Deformation Behaviors of a Tungsten Penetrator (A) and Alpha Uranium Alloy Penetrator (B).

The criteria for the development of shear instability can be expressed in equation form below:

$$\frac{d\tau}{d\gamma} = \left.\frac{\partial\tau}{\partial\gamma}\right|_{\theta,\dot{\gamma}} + \left.\frac{\partial\tau}{\partial\dot{\gamma}}\right|_{\theta,\gamma}\frac{\partial\dot{\gamma}}{\partial\gamma} + \left.\frac{\partial\tau}{\partial\Theta}\right|_{\gamma,\dot{\gamma}}\frac{\partial\Theta}{\partial\gamma} = 0 \tag{1}$$

(change in   = (strain   + (strain-rate + (thermal
 shear strength)   hardening)   hardening)   softening)
           where τ = the shear strength of the material,
                  γ = plastic shear strain,
                  $\dot{\gamma}$ = shear strain rate,
           and Θ = temperature

A quantitative measure of a material's susceptibility to adiabatic shear failure can be obtained by solving Eqn.(1) for a critical shear strain, γ_c , at which the thermal-softening of the material overcomes its strain and strain-rate hardening, i.e. when $d\tau/d\gamma = 0$. The predicted values of the critical strains will depend on the particular functional relationships, linear, power law, exponential, etc. assumed for each term in Eqn.(1).

Ignoring thermal effects, most materials, even nominally brittle materials like pure tungsten, will deform and back-extrude in a ductile manner under the high hydrostatic pressures operating at the penetrator-target interface. When the heat being generated by the rapid deformation and the resulting rise in temperature are considered, however, the material's deformation behavior, uniform bulk plastic strain or localized deformation, will be determined by the balance of strain-hardening (strain, strain-rate) and strain-softening (thermal) effects expressed in Eqn.(1). In materials with low thermal-softening rates, such as tungsten or tantalum, shear instabilities will not develop and their deformation will resemble the illustration in Figure 8a or in Figure 6a. The mushroomed head on the penetrator grows quite large before the eroded material is discarded, and the resulting penetration performance suffers. By contrast, the alpha phase uranium alloys rapidly lose strength between room temperature and the temperature of its first phase transformation of approximately 600° C. As a result of uranium's high rate of thermal-softening, the deformation is quickly localized in adiabatic shear bands. For this reason, little plastic deformation is observed within each of the shear chips (Figure 5c). The mushrooming of the penetrator is cut off periodically, as the deformation reaches a critical plastic strain level and is then relieved by the shear failures (Figure 8b). The roughened penetration tunnel walls result from this periodic process. The average size of the head on the penetrator, and the volume of armor material which must be displaced by the moving penetrator-target interface with each increment of penetration, are reduced. The erosion of the rod by localized failures, instead of by uniform bulk deformation, may further improve penetration by also reducing the kinetic energy expended to deform the penetrator itself.

The deformation behavior and resulting ballistic performance of the nickel-base matrix WH alloys is intermediate between the two extremes exhibited by the pure tungsten or W-Ta composite, and the alpha uraniums. The conventional WH alloys undergo extensive bulk deformation (Figure 4c) like the W-Ta composite, but eventually a localization of the deformation and discard of the material as discrete chips occurs. Clearly, the melting temperature of the nickel-base alloy (approximately 1500° C) being lower than that of tungsten or tantalum, raises the thermal-softening rate in the matrix, and aids the localization. The oriented microstructure (Figure 4b-c), which evolves as a result of the deformation, may also be necessary for the shear bands to finally propagate in the matrix. The mushroomed heads which develop on the penetrators prior to the discard, however, are quite large (Figures 4a-b), and the ballistic performance of the nickel-base matrix WH alloys though a slight improvement over the W-Ta alloy, are far poorer than comparable density alpha uraniums. Replacing the nickel-base matrix with an alloy that is more prone to adiabatic shear banding, rather than more resistant (as was the tantalum), might induce these localizations in the overall composite, earlier in the deformation process. The ballistic performance of these shearing WH alloys might be significantly improved. This approach is now being pursued in several BRL-sponsored programs, and among the matrix alloys being investigated are steel, cobalt, titanium and hafnium. Although the deformation behaviors and ballistic performances of the different materials seem to correlate well with

their thermo-mechanical properties, the possible contribution of pressure-shear transformations must also be considered in these future studies.

An explanation for the differing ballistic responses of the DU and WH alloys to changes in mechanical properties also derives naturally from the adiabatic shear interpretation. Increasing the strength of an alloy increases the work done, and therefore the heat generated, with each increment of plastic strain. The rise in temperature with plastic strain ($\partial \Theta / \partial \gamma$) is therefore greater in the higher strength alloys and the instability condition expressed in Eqn.(1) will develop more quickly. The critical strains required to initiate the adiabatic shear failure will decrease with increasing alloy strength. This relationship between strength and critical strain has been observed in both steel and titanium alloys[16], as well as for DU alloys. Increasing the strength of the uranium alloy therefore improves its ballistic performance by further reducing the amount of plastic strain (mushrooming) which must occur before the shear failures develop. Likewise, decreasing the DU alloy's strength delays the discard of the shear chips and its performance is degraded. The extreme example in this direction is the monoclinic U-6Nb alloy. Its low initial strength, perhaps aided by changes in the thermal-softening rate due to shifts in phase transformation temperatures, eliminated the adiabatic shear failures entirely. Its ballistic performance was therefore much worse than those of the alpha uraniums.

From the heavily deformed microstructures within the discarded chips (Figure 4c), it is clear that the critical strains are much larger for the WH alloys. With this degree of plastic deformation being applied to the alloys, initial differences in the strengths of the WH alloys are quickly obscured, and their net effect on critical strain values became quite small. As a result, the apparent influence of WH alloy strength on ballistic performance is much less than that for the alpha uraniums.

5.0 Conclusions

The ballistic testing and microstructural analysis of the performance of uranium and tungsten heavy alloys indicate that a thermo-mechanical (adiabatic shear) instability is responsible for uranium's superior performance. During the penetration of armor, a WH alloy projectile undergoes extensive bulk plastic deformation, building up a large mushroomed head on the rod before the penetrator material is discarded. In the case of a DU alloy projectile, however, localized adiabatic shear failures periodically cut-off the build-up of the mushroomed head. This minimizes the volume of target material displaced by the moving penetrator-target interface and allows the DU projectile to expend its kinetic energy making a deeper, rather than wider, tunnel into the armor.

Prior tungsten alloy development programs have failed to achieve any significant improvements in the fundamental penetration performance of WH alloys. These efforts have centered on improving engineering properties in the hopes of increasing a penetrator's

resistance to deformation. The analysis above suggests an alternate approach, that is conceding the failure of the penetrator material, but controlling the manner in which it fails. One approach, already outlined above, is to induce a thermo-mechanical instability in the WH alloy by replacing the nickel-base matrix alloy with one more susceptible to adiabatic shear localizations. An alternate approach, which does not rely on thermal-softening, is to create similar localizations of the deformation by structural or microstructural design. This might include prior orienting of the microstructure or the use of preformed shear chips. All of these techniques, under the broader heading of alterations of deformation behavior, are now being studied at the BRL, and offer the potential of achieving significantly improved penetration performance with WH alloy projectiles.

Acknowledgment

REFERENCES

1. Nicolas, T., (1980). Dynamic Tensile Testing of Structural Materials Using a Split Hopkinson Bar Apparatus, Air Force Wright Aeronautical Laboratories Techn. Rep. AFWAL-TR-80-4053.

2. Johnson, G. & Cook, W., A Constitutive Model and Data For Metals Subjected to Large Strains, High Strain Rates, and High Temperatures, Proc. of the 7th Intnl. Symp. on Ballistics, The Netherlands, April 1983.

3. Tate, A. (1967). A Theory for the Deceleration of Long Rods After Impact, J. Mech. Phys. Solids, 15, 387.

4. Frank, K. & Zook, J. (1986). Energy-Efficient Penetration and Perforation of Targets in the Hypervelocity Regime, Proceedings of the 1986 Hypervelocity Impact Symposium, International Journal of Impact Engineering, V.5, No.1-4, 277-284.

5. Zukas, J.A. et al.,(1981). Three-Dimensional Impact Simulations: Resources and Results. Computer Analysis of Large-Scale Structures, AMD, V. 49, American Soc. of Mechn. Eng.

6. Bloore, E. & Herr, L., (1978). Penetrator Materials Development (U). Published in the Agenda of Presentations Given at the 25-28 April 1978 US-FRG Meeting Asso. with DEA-G-1060 Ballistic Research and Development, V.2, Log No. RJE(SP)-5-S-78.

7. Magness, L., (1988). Materials For Kinetic Energy Projectile Applications, Proceedings from the 28th NATO DRG Seminar on Novel Materials for Impact Loading, Bremen, Germany. (Unclassified).8. Magness, L., Farrand, T. & Van Rensselear, N., (1986). Evaluation of Depleted Uranium Alloys For Use in the XM881, 25mm ADFSDS-T Cartridge (U), BRL Mem. Rep. BRL-MR-3563 (Confidential).

9. Hantel, L.W. & Taylor, J.W. (1974). PHERMEX Evaluation of Air Force Tungsten and U-0.75% Ti Penetrators (U), Los Alamos National Lab. Rep., LA-5658. (Unclassified)

10. Sandstrom, D., Dunn, P., & Hogan, W. (1986). Comparison of Tungsten and Uranium Kinetic Energy Penetrators Fired Into Semi-Infinite Steel Targets, Proceedings of Tungsten Ordnance Technology Seminar, Sept. 29-30, 1986, Washington, DC.

11. Foster, E. et al, (1987). Preparation of Ballistic Test Specimens From Tantalum-Coated Tungsten Powders, Battelle, Columbus Division, Contract Rep. to BRL.

12. Fulton, F.J., Cline, C.F. & Snell, E.O., (1980). Penetration of Mild Steel Targets By High-Density Long-Rod Penetrators (U), Lawrence Livermore Lab. Rep., UCRL-52991. (Confidential).

13. Gerlach, U., (1986). Microstructural Analysis of Residual Projectiles - A New Method to Explain Penetration Mechanisms, Metall. Trans., V.17A, 435-442.

14. Irwin, G.J., (1972). Ballistic Evaluation of Materials For Armour Penetrators (U), Defense Researche Establishement Valcartier Rep., DREV-R-643/71, (Unclass./Limited)

15. Rogers, H., (1983). A Review of Adiabatic Shear, in Material Behavior Under High Stress and Ultrahigh Loading Rates, eds. Mescall & Weiss, Plenum Press, N.Y., 110-118.

16. Staker, M.R., (1981), The Relation Between Adiabatic Shear Instability Strain and Material Properties, Acta. Met.,V.29,683-689.

Thin Polymers with Novel Dielectric Properties

* Robert J. Mammone, Dr.
Michael Binder, Dr.
U.S. Army Electronics Technology and Devices Laboratory,
LABCOM, Power Sources Division
Fort Monmouth, NJ 07703-5000

INTRODUCTION

The U.S. Army has numerous critical applications which call for very large quantities of electrical energy to be stored and then discharged in submicrosecond time intervals in order to power a variety of devices. Such energy storage and energy delivery requirements can be most efficiently achieved by using capacitors to store electrostatic charge. The most basic capacitor consists of two parallel conducting plates separated by vacuum. The vacuum acts as the insulating dielectric which stores energy. When the two conducting plates are separated not by vacuum but by a thin layer of insulating dielectric material, the energy density or amount of electrical energy that can be stored is now increased and given by the equation:

$$\text{Energy Density} = \tfrac{1}{2}\,\epsilon\epsilon_o AV^2/D \qquad (1)$$

Here, ϵ_o is the permittivity of free space, 8.85×10^{-14} F/cm, the product $\epsilon\epsilon_o$, is the relative permittivity of the material relative to the permittivity of vacuum, while ϵ is commonly referred to as the dielectric constant. The dielectric constant, the factor by which the energy storage of the chosen dielectric exceeds that of vacuum, is not a constant but is highly frequency and temperature dependent. As a rough guide, ϵ for a perfect vacuum is 1 and ∞ for a perfect conductor. A is the effective surface area of the dielectric sandwiched between the two plates, and V, is the voltage applied across the dielectric of thickness D. For practical pulsed power capacitor systems, achieving the highest energy and power densities under high reliability conditions are the most critical considerations. Ultimately, it is hoped to develop storage devices with battery like capacity and with capacitor like discharge characteristics.

Most organic solids, especially commercially familiar polymers, are insulators with low dielectric constants. Innovative approaches are required in order to develop dielectric materials having high dielectric constants and high voltage breakdown strengths for next generation dielectrics in high energy density storage capacitors. By studying the variables in Equation 1, one can determine the important research directions that must be taken in order to increase maximum energy storage capabilities. In equation 1, the ratio, A/D, is an engineering and design variable which is basically determined by the capacitor size used and manufacturing techniques chosen. The two remaining terms, the relative dielectric constant, ϵ, and operating voltage, V, are somewhat under control of material scientists and will be investigated here. Since dielectric losses are required to be small, one would prefer that the bulk material be a good insulator.

Obviously, because energy density increases as the square of V, to achieve the highest energy density, one would normally operate at the highest possible value of V. The maximum voltage that can be tolerated by a given dielectric film before it breaks down and conducts is called its breakdown strength, usually given in units of Volts/mil. The intrinsic breakdown strength depends on the number of mobile electrons in the dielectric insulator and upon their energy distribution. It is limited by the bond strength of the weakest bond. Breakdown strength is depressed by the presence of ionic impurities voids or absorbed moisture. Unfortunately, there are relatively few reliable methods for theoretically predicting what the breakdown strengths of materials will be. In addition, the number of charge and discharge cycles that the dielectric can tolerate without becoming destroyed decreases as V^5. This means that although a dielectric operating at high voltages will have very high energy storage capabilities, it will also not be useful for many cycles. We therefore concentrated our research efforts on developing methods of forming new materials having increased values of ϵ. Unfortunately, in nature, the two variables C and V are inversely related. Commercially available solid polymeric materials such as polycarbonate and polypropylene have high breakdown strengths (6-10KV/mil) but also low dielectric constants (3-4) while ceramics such as barium titanate having high dielectric constants (750) tend to have low breakdown strengths (100V/mil).

In predicting dielectric constants, considerable efforts (both theoretical and experimental) have been made to establish relationships between chemical structure of organic molecules and their efficiencies as capacitor dielectrics. In this regard, intrinsic bulk dielectric properties of conductor/insulator composites has recently received much theoretical and experimental attention. One theory, Percolation Theory, has predicted that bulk dielectric constants (and conductivities) of solids will abruptly increase when a bulk insulator is uniformly doped with a small percentage of conductive components (1-7). This sort of

abrupt compositional induced changes were found to be true experimentally i.e. effective dielectric constants exhibited dramatic increases in many solid systems such as carbon particles embedded in bulk insulators. In those cases, theoretical studies have predicted that effective bulk dielectric constants should increase with concentration of conductive dopant in a power law relationship:

$$\epsilon = \mathcal{K}(P_c - P)^{-s} \qquad (2)$$

Where P is the volume percent of conductor in the bulk insulating matrix, P_c is the percolation threshold or the critical concentration of conductor (the point where the dielectric constant of the bulk conductor insulator mixture increases dramatically as P approaches P_c from below), and s is a critical exponent typically less than unity.

The physical origins of this dramatic increase of dielectric constants are not exactly clear. It is believed to be due to the fact that in heterogeneous conductor-insulator mixtures having two distinct phases conducting domains are separated from each other by insulating barriers. Since there is no communication between the localized conducting regions (charges are separated and cannot pass through insulating regions), electric charges accumulate at the conductor insulator interface. Development of this interfacial charge polarization within the bulk material contribute to huge dielectric constants and abnormally large effective capacitance since the interfaces are all connected in parallel. As the concentration of these conducing centers increases near the percolation threshold, the large numbers of conductive regions, each separated by very thin insulating barriers lead to huge bulk dielectric constants.

Our research goal was to increase effective bulk dielectric constants (and therefore capacitance) of thin, easily formed, flexible polymer films that could be used as solid dielectrics in wound capacitor designs. In order to minimize dielectric losses, the bulk material would have to be highly insulating. By starting with good insulators having low losses, bulk dielectric constants can now be hopefully increased by physically dispersing small quantities of conductive regions within the bulk insulating matrix. We attempted to recreate the above scenario in a number of solid polymeric systems by utilizing percolation/effective medium theory to design and form suitable bulk insulators that have very high dielectric constants. These materials are made by embedding or dispersing various types of conducting entities within bulk insulating films. Thin films of these new materials were then characterized in terms of dielectric properties including dielectric constants, dielectric or breakdown strengths, resistivity and dielectric losses. [Dielectric loss is the power dissipated in dielectrics due to conductive processes]. Success of our molecular

design is rated by the ability to create new kinds of materials having desired dielectric responses. By careful engineering of selected molecular properties we hoped to achieve the desired macroscopic properties. We shall discuss three such examples taken from our ongoing in-house research and also outline a new approach that we believe has future application in creating materials having extremely large bulk dielectric constants.

1) ELECTROCHEMICAL DOPING: In electrochemical doping of conjugated polymers with anions, the anions themselves form positively charged sites on the polymer backbone. In this case, the positively charged backbone chain and incorporated anion serves as the conductive center. Increased values for bulk dielectric constants have been observed.

2) ION IMPLANTATION: Creation of damage regions within bulk selected insulating polymers by bombardment with high energy ionizing radiation. Here, localized damaged regions act as conductive domains. Enhanced values for bulk dielectric constants have been observed.

3) POLYMER COMPOSITES: Formation of polymer composites by blending conducting and insulating polymers is an effective means of tailoring materials to achieve a specific balance of properties. Enhanced values for bulk dielectric constants have been observed.

Improvements in overall dielectric properties are expected in all of the above two-component systems since conductive species furnish the high dielectric constant while the host insulator furnishes the high breakdown strengths, bulk insulation, and other desirable mechanical properties.

1. ELECTROCHEMICAL DOPING

Easy, one step electrochemical synthesis of compact, organic polymer films deposited on metal surfaces has made electrically conducting polymers very popular. The advantage of electrochemical polymerization is that not only can polymers be prepared with virtually any desired thickness on various surfaces, but their specific conductivity can be controllably varied over many orders of magnitude. It is well known that electrochemically synthesized conjugated polymers can be transformed into highly conducting polymers by redox reactions which, by analogy with semiconductor terminology, are referred to as doping. It is not, however, the same phenomena as that which occurs with traditional inorganic semiconductor like silicon or gallium arsenide. When polymers are electrochemically oxidized and thereby made conductive, the process involves removal of electrons from the polymer chain. The polymer, left with a net positive charge, is then

neutralized or balanced by a simultaneous insertion of a dopant counter anion from the electrolyte solution into the polymer chain. Doping essentially involves formation of a polymer-anion complex that preserves electroneutrality. In certain systems the inserted anion is present in concentrations as high as one counterion for every three or four repeat monomer units.

In electrochemical polymerization, the polymer properties depends on a number of variables including the electrode material, its surface, the electrolyte, counterions and temperature used. We wanted to study electropolymerized polymers that has <u>low</u> bulk conductivities (i.e. were highly resistive) even when fully doped. "Metallic islands" consisting of a positively charged backbone chain and incorporated dopant anion should produce a polymer with a high dielectric constant.

As a model system, we investigated one polymer taken from the polypyrrole family of polymers. As shown in Figure 1 a pyrrole monomer is a ring of five atoms of carbon and one nitrogen. Polypyrrole can be viewed as a linear, planar polymer in which nitrogen atoms

FIGURE 1: THE STRUCTURE OF POLY-N-METHYLPYRROLE
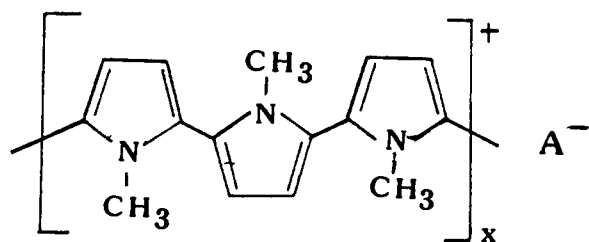
of the monomers along the chain alternately point in opposite directions. This limits any steric interactions and allows for the highest overlap of orbitals in the conjugated system. However, in the case of poly n-methylpyrrole, PNMP, substitution of the nitrogen bonded hydrogen with a methyl group forces the chain out of planarity by twisting monomer units relative to one another. Chain distortion decreases the mobility of the charge carriers on the chain, inhibits interchain interactions and ultimately causes a decrease in the conductivity of the polymer obtained. This accounts for the approximately 10^4 decrease in conductivity in going from polypyrrole to poly n-methylpyrrole. We therefore studied PNMP as a model compound since its decreased conductivity made it of interest as a dielectric in capacitor applications.

We have found that the choice of supporting electrolyte, anions in the supporting electrolyte and temperature of polymerization all have a

great influence on electrical and mechanical properties of electrochemically formed PNMP (8). We varied polymerization parameters and established the conditions necessary to form films that have application in capacitors.

Surprisingly, when PNMP is formed in electrolyte containing sulfate or dihydrogen phosphate anions, the resulting highly <u>insulating</u>. This is in sharp contrast to PNMP formed in the presence of organic anions such as hydroxybenzenesulfonate anions which is <u>conductive</u>. At this point we are not entirely sure why sulfate or dihydrogen phosphate anions produce an insulating polymer. We should point out that when other polymers such as polypyrrole are electrochemically formed in electrolytes containing either of these two anions, the resulting polymer is highly <u>conductive</u>. It is only the combination of sulfate or dihydrogen phosphate anions in the supporting electrolyte and n-methylpyrrole that yields an electrochemically formed <u>insulating</u> PNMP.

TABLE 1

Dielectric Properties of Poly-N-Methylpyrrole Prepared from Aqueous Solutions Containing Sulfate, Dihydrogen Phosphate and 4-Hydroxybenzenesulfonate Anions at R.T. and 75°C.

ANION	TEMP	CONDUCTIVITY S/cm	DIELECTRIC CONSTANT @10KHz	TAN DELTA @10KHz	BREAKDOWN VOLTAGE V/mil
Dihydrogen Phosphate	RT	6.3×10^{-9}	7.0	0.602	1120
Sulfate	RT	4.0×10^{-10}	8.4	0.137	980
4-Hydroxy Benzene Sulfonate	RT	5.2×10^{-4}	----	>10	---
Dihydrogen Phosphate	75°C	3.0×10^{-12}	3.8	0.054	1350
Sulfate	75°C	6.0×10^{-12}	6.6	0.045	650
4-Hydroxy Benzene Sulfonate	75°C	3.6×10^{-6}	30.8	>10	---

2. ION IMPLANTATION

Poly-p-xylylenes, available commercially from Union Carbide under the trade name of Parylene, are well known conformal vapor deposited coatings that are primarily used as insulators in electronics. These films would have applications as dielectrics in capacitor applications were it not for their low dielectric constant. We were successful in increasing the dielectric constant of these thin films through ion bombardment (9).

We bombarded 1.6 micron Parylene films (coated on aluminum coupons) with 50 and 150 keV chloride and argon ion beams and varied the flux density (the number of ions impinging on the sample) from 10^{14} to $2x10^{16}$ ions/cm^2. Ion energies chosen were designed to be large enough so impinging ions could create damage centers deep in the bulk of the film. Thus, unlike ion implantation in thick solids, where only surface regions are affected, in our case, ions embedded in the bulk were modifying bulk dielectric properties of the material.

Our results shown in Figure 2 and summarized in Table 2 indicates that dielectric properties of Parylene film is a strong function of ion dose. Dielectric constants and resistivities of irradiated Parylene films remained relatively constant for a wide rang of ion flux and then showed a sudden and dramatic change as the ion flux for either chloride or argon ions rose beyond $2x10^{15}$. Dielectric constants of Parylene film rose from a value close to 3 for unirradiated film to approximately 15-20 for films irradiated with $1x10^{16}$ ions/cm^2. These large measured dielectric constants did not change even after samples were allowed to anneal overnight. For ion doses greater than 10^{16} ions/cm^2, dielectric constants decreased.

Figure 2: Dielectric constant (■) and Resistivity (o) of Parylene for various doses of argon ions as a function of ion flux.
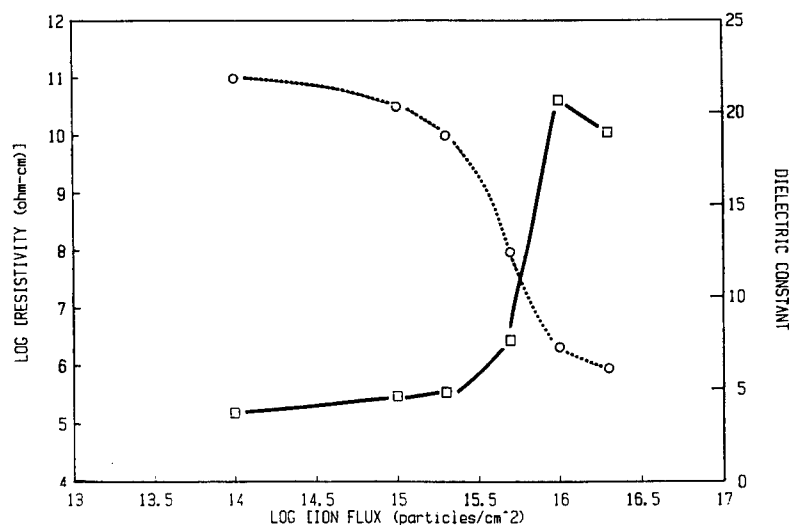
TABLE 2

Measured dielectric properties for Parylene C film 1.6 um thick before
and after various doses of chloride and argon ion implantation.

Ion fluence	Film color	Dielectric constant	tan δ	Resistivity (through film) (Ω cm)
None	Clear	2.82	0.0381	2.4×10^{11}
1×10^{14}	Pale yellow	3.15 c	0.0266 c	1.6×10^{11} c
		3.70 a	0.0567 a	9.8×10^{10} a
1×10^{15}	Gold	3.33 c	0.492 c	7.3×10^{10} c
		4.60 a	0.061 a	3.2×10^{10} a
2×10^{15}	Brown	4.02 c	0.510 c	5.7×10^{10} c
		4.81 a	0.083 a	1.0×10^{10} a
5×10^{15}	Dark brown	5.20 c	1.183 c	2.7×10^{7} c
		7.63 a	0.441 a	9.5×10^{7} a
1×10^{16}	Dark brown–black	14.0 c	2.365 c	6.9×10^{6} c
		20.7 a	4.483 a	2.1×10^{6} a
2×10^{16}	Dark brown–black	9.59 c	4.56 c	4.0×10^{6} c
		18.9 a	11.3 a	9.0×10^{5} a

Surface area of the contacts was 1.27 cm². All values were measured at room temperature
and 10 kHz. The letters c and a next to the various dielectric values refer to chloride and
argon ion irradiation respectively.

These observed dielectric changes can perhaps be explained as
follows: As ion fluence rises, localized damage centers (which are
conductive), are created inside the bulk insulating material. These
aggragates, at first isolated, primarily at low ion doses, grow in
number with increased ion dose until they begin to fuse together. It is
interesting to note that dielectric constant of Parylene films after
argon ion irradiation is consistenly higher than that for comparable
doses of chloride ions. This may be due to increased damage of the
larger sized argon ion.

3. POLYMER COMPOSITES
We have attempted to form homogeneously blended polymer composites
consisting of materials having electronic conductivity dispersed in an
insulating host polymer. We focused our attention on polymers or
oligomers that can be film cast or molded to produce thin high
dielectric constant films. As our host polymer we chose the widely
investigated piezoelectric and ferroelectric Poly vinylidene fluoride,
PVDF, since it exhibits a variety of scientifically interesting and
technologically important properties.

Figure 3: Structure of POMA (a) and PVDF (b).

(a)

(b)
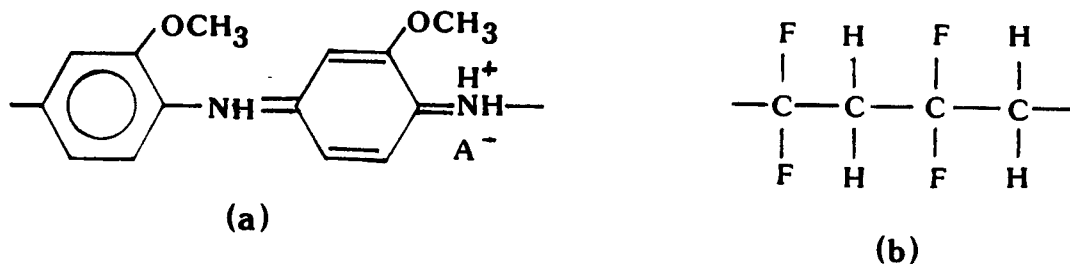
PVDF with a backbone of CH_2-CF_2 repeat units possesses an an unusually high value of 8-11 for its relative dielectric constant, relatively low dissipation factors and high electric field breakdown strengths. These properties, together with its excellent thermal and chemical stability have made PVDF interesting as a useful solid dielectric for conventional capacitor and piezoelectric applications (hydrophones). We chose poly o-methoxyaniline, POMA, as the conductive dopant since both POMA and PVDF are soluble in dimethylformamide and thus freestanding and flexible films are easily cast from solution. We wanted to combine desired properties of each component namely the high dielectric constant of the conductive polymer with high dielectric strength and low lossess of PVDF matrix.

Casting techniques, widely used for film production, is actually a coating process in which a polymeric film is deposited on a selected substrate. In casting processes the polymer dissolved in a suitable solvent is placed on a flat substrate. The substrate is moved at constant speed and the polymer solution is forced under an adjustable knife blade. The gap between the blade and the substrate determines the coating wet thickness and ultimately the thickness of the resulting cast film. Evaporation of the solvent causes a uniform reduction in film thickness by an amount which depends on the % solids in the solution. After drying, the formed polymer film can be removed from the substrate.

We have performed extensive dielectric measurements on these films and characterized how dispersed polymers percolate. Dielectric properties for solution cast composites of PVDF films containing various percentages of POMA have been measured over the frequency range of 20 Hz up to 100 kHz. Large increases in dielectric constants (up to a factor of six at 1 kHz) were observed in PVDF films containing approximately 20% POMA.

Figure 4: Dielectric constant of PVDF/POMA composites for various %
POMA at 1 KHz.
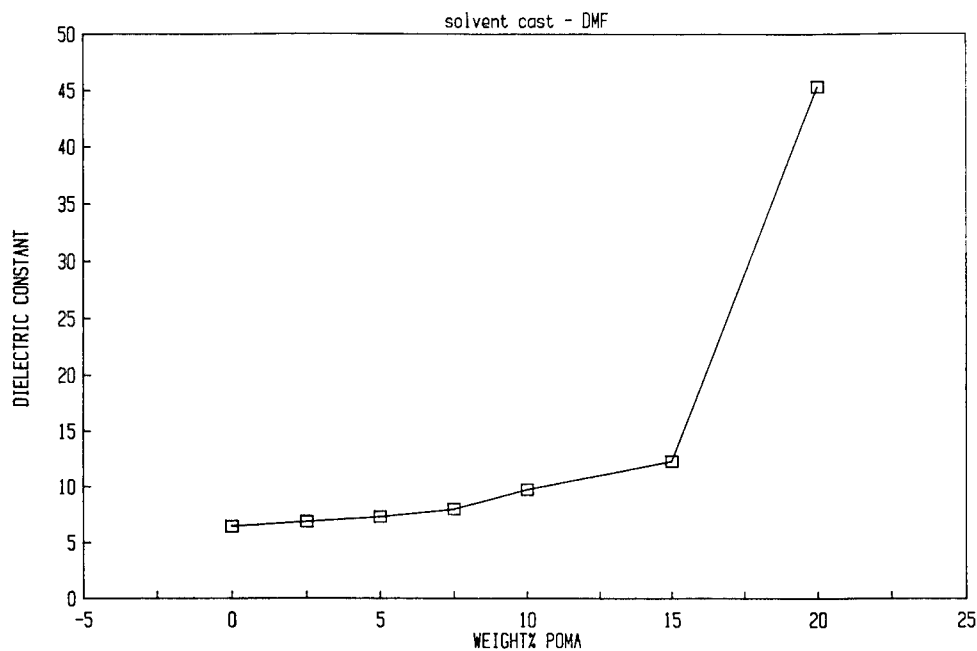
PVDF/POMA composites represent a good example of results obtainable
near the percolation threshold in polymer composites. Our results
indicate that it is possible to produce flexible PVDF composites with
dielectric constants significantly greater than those of pure PVDF.

4. FUTURE EXPERIMENTS

Polymers containing highly conjugated delocalized pi electron
structures have generated considerable interest as conductive materials.
Recently, there has been heightened interest in these materials because
of their nonlinear optical properties. Non centrosymmetric structures
can be prepared by replacing atoms with various substituents which may
act as electron donors or acceptors. The added substituents distort the
pi electron cloud which gives rise to asymmetric pertubation of electron
distribution under the influence of electric fields. These polymers are
attractive for our applications since they can produce large electronic
effects (distortion of electronic distribution about the nuclei in an
electric field) which can yield strikingly high values for dielecctric
constants.

ACKNOWLEDGEMENTS

REFERENCES

1. H. Scher and R. Zallen, J. Chem. Phys. $\underline{53}$, 3759 (1970)
2. D.J. Bergman and Y. Imry, Phys. Rev. Letters, $\underline{39}$, 1222 (1977)
3. D.J. Bergman, Phys. Rev. Letters, $\underline{44}$, 1285 (1980)
4. D.M. Grannan, J.C. Garland, and D.B. Tanner, Phys. Rev. Letters, $\underline{46}$, 375 (1981)
5. Y. Gefen, A. Aharony and S. Alexander, Phys. Rev. Letters, $\underline{50}$, 77 (1983)
6. Y.C. Chou and T.S. Jaw, Solid State Comm., $\underline{67}$, 753 (1988)
7. Y. Song, T.W. Noh, S. Lee and J.R. Gaines, Phys. Rev. B., $\underline{33}$, 904 (1986)
8. R.J. Mammone and M. Binder, to appear in J. Electrochem. Soc.
9. M. Binder and R.J. Mammone, to appear in Surface and Coatings Technology

491

Research and Development by USACE to Enhance
Logistics-Over-The-Shore (LOTS) Operations (U)

MAJ James N. Marino and Dr. Lyndell Z. Hales
USAE Waterways Experiment Station
Coastal Engineering Research Center
3909 Halls Ferry Road
Vicksburg, Mississippi 39180-6199

Present and Increasing Requirement for LOTS Capability

As expressed by US Transportation Command (USTRANSCOM) (1990) ("Joint Logistics Over the Shore (JLOTS) III, Feasibility Study") "...in the future, as in the past, 95 percent of the supplies and equipment required by deployed operating military forces will be transported via strategic sealift...The potential need for conducting JLOTS operations is increasing in the face of changing relationships worldwide. Recent events in Eastern Europe hold the potential for unprecedented changes in relationships between nations which may lead to reductions in overseas force structure. This would place increased emphasis on the transportation infrastructure, including JLOTS capability, needed to support the return of forces and sustaining operations. At the same time, the likelihood of military operations in Third World locations increases. Most Third World countries lack the fixed port infrastructure to support significant quantities of resupply shipping. Thus, every indicator points to the increased need for a robust JLOTS capability..."

Evolution of LOTS Research and Development

The genesis of the required capability for Logistics Over the Shore (LOTS) operations was a 2-year comprehensive study by Defense Logistics Review Board (1970) ("Logistics Support in the Vietnam Era, Monograph 18, Transportation and Movement Control,"). That study assessed experience and likely military and political situations where it would be necessary for US forces to enter upon a land mass without the benefit of well-established and equipped sea-ports. The state of US military and merchant shipping to support logistics operations in major contingencies, also, was assessed. The Board's conclusions and recommendations clearly highlighted the need to alter the Defense logistics system to provide an adequate over-the-shore support capability, and to bring the system in line with trends by the shipping industry to enable full use of available assets when required.

As follow-up action, the Office, Secretary of Defense (OSD) assigned responsibility to the Assistant Secretary of Defense (ASD), Manpower, Reserve Affairs, and Logistics (MRA&L) to develop and pursue the necessary programs to achieve a capability to conduct LOTS operations. With advice from the Army, Navy, and Marine Corps, ASD MRA&L published a Department of Defense (DOD) Project Master Plan (PMP) in January 1973 which assigned responsibilities to the various Services for specific segments of the total effort. Responsibility of Department of the Army (DA) in this regard was directed primarily toward development of specialized pieces of equipment for movement of containers, supplies, and oversized paraphernalia after such items had been placed onshore from supply ships anchored offshore (with the exception of the Army's Temporary Container Discharge Facility (TCDF) (a modified B-DeLong barge). The original and updated versions of the PMP keynote the need to use merchant shipping to augment US Fleet sealift capabilities in major military contingencies. The PMP also notes the shipping industry's greatly increased use of standardized freight containers. To accommodate containerized shipments, ocean carriers are replacing breakbulk cargo vessels with containerships at a steady rate. The containerships consist mainly of non-self-sustaining types, whereby ship onloading and offloading are accomplished at modern port terminals with deep-draft piers and gantry cranes fitted with quick-locking container lifting frames or spreaders.

In recognition of these activities, ASD MRA&L took action to acquire a containerized distribution system within the Defense logistics structure to enable employment of commercial containerships in contingencies, as well as a means to discharge these ships at their offshore anchorage when port facilities are not available in the military objective area. Over the Shore Discharge of Containerships (OSDOC) was formulated in a system definition paper promulgated by ASD MRA&L (1975) and incorporated into the PMP. The Navy was assigned primary responsibility for developing cargo offloading and transfer capabilities to satisfy the stated requirement. Subsequently, the Chief of Naval Operations (OPNAV) announced an operational requirement for a Container Offloading and Transfer System (COTS), and tasked the Navy Materiel Command (NAVMAT) with the development of COTS. In turn, NAVMAT designated the Naval Facilities Engineering Command (NAVFAC) as the development agency for COTS and other efforts to satisfy the Navy's development assignments. OPNAV then transferred responsibility for developing a containership offloading capability to the Naval Sea Systems Command (NAVSEA), which initiated the development of the Tactical Auxiliary Crane Ship (TACS).

DOD and Navy concluded that the investment to obtain a LOTS capability with full reliance on military shipping was unrealizable with respect to peacetime manpower and inventories that would have to be maintained. A military solution consistent with operational and economic factors was needed. Cargo discharge facilities must be available when required for container, tanker, and other ships offshore in advanced operational areas. These facilities must be cost effective and retrievable when no longer needed to support logistics operations. An integrated and balanced cargo handling capability, including ship-to-shore transfer, must be developed for the full spectrum of contingency logistics operations associated with

amphibious operations (both Assault Echelons (AE) and Assault Follow-On Echelons (AFOE)), and sustainment support. Attendant basic requirements provide for the transfer of cargo from containerships to lighterage units for transfer to shore, and for the retrograde of empty containers or containers with cargo being evacuated from the objective area. The turnaround of container assets is an important aspect of the complete spectrum of logistics support.

OPNAV devised the Sealift Support Facilities Program (SSFP) to provide the means to conduct logistics operations over the shore. It is one element of the Strategic Sealift Program (SSP) which is concerned with the transportation of forces and their logistics support to meet military threats. The SSFP is designed to furnish the resources necessary to establish the interface between seaborne logistics operations and logistics operations ashore within specified time frames in a non-hostile environment. Its primary goal is to meet requirements for the ship-to-shore transfer of large volumes of dry and bulk liquid cargo in support of Fleet Marine Forces (FMF) engaged in amphibious operations. A primary requirement of SSFP is to provide maximum commonality and compatibility of components that can be used by Army, Navy, or Marine Corps units with military and civilian personnel.

Army/Navy Cooperative Logistics Efforts

In 1982 a Memorandum of Agreement (MOA) was signed by Army and Navy headquarters logistics representatives to study and clarify joint inter-service responsibility in the development, acquisition, and use of Army/Navy Cargo Offload and Discharge Systems (COLDS). Subsequently, extensive coordination and cooperation achieved highly effective results in logistics planning, OSD-sponsored mobility studies, mobility program balance, service support for a balanced program, and mutually supportive programs and objectives in support of Joint Chiefs of Staff's (JCS) and Commander-in-Chief's (CINC) mobility requirements. The service chiefs endorsed the cooperative logistics effort, and affirmed the need for continuity and expansion of that effort. It was decided to continue the appointment of an Army liaison officer to the Navy Materiel Command Strategic Sealift Support Office to coordinate technical matters concerning watercraft and other aspects of Strategic Sealift interface between Army and Navy. It also was decided to prepare the Joint Logistics Over the Shore II (JLOTS II) exercise and demonstration scheduled for late 1984 to cover the full range of events and equipment, with particular emphasis on throughput of supplies and equipment across the surf zone.

Service Regulations

Departments of the Army, Navy, and Air Force regulations of 1 June 1983 (Army Regulation No. 55-176, OPNAV Instruction 4620.8B, Air Force Regulation No. 75-4, Marine Corps Order 4620.6) ("Logistics Over the Shore Operations in Overseas Areas") defines Service responsibilities, and applies to all Active and Reserve Components involved in LOTS operations in overseas areas. A logistics operation conducted in support of the AE or AFOE phase of an amphibious operation in a hostile, or potentially

hostile environment is termed an Amphibious Logistics System (ALS). A logistics operation conducted in friendly territory or in a nonhostile environment, when it is anticipated that there will be no opposition from enemy forces other than covert operations by small teams or irregular forces, is considered a LOTS operation. Such a LOTS operation will normally be conducted by the Army as part of base, garrison, or theater development operation which immediately follows an amphibious or joint amphibious operation.

LOTS operations will be conducted over unimproved shorelines, through fixed ports partially destroyed by combat action, through shallow-draft ports not accessible to deep-draft shipping, and through fixed ports that are inadequate without utilizing LOTS capabilities. LOTS operations consist of loading and unloading breakbulk, roll-on/roll-off vehicles, containers, and bulk petroleum-oils-liquids (POL) and water cargo from ships in the theater of operations. LOTS activities include operation of ships, watercraft, and lighterage in the loading and unloading areas. Sustained throughput operations over unimproved shorelines will continue through a Sea State 3 condition (waves up to 5 ft high with periods up to 7 seconds, 30-knot winds, 4-knot currents, 7-ft surf, and a tidal range of 8 ft) onto open beach areas with gradients of 1-to-50. LOTS sites will be selected by agreement between the Navy Commander and the Army LOTS Commander, and will be approved by the Unified Commander (UC).

JLOTS II Exercises

The largest peacetime LOTS exercise ever conducted by US military forces, JLOTS II, was conducted at Fort Story, Virginia, in September-October 1984. Over 3,000 personnel from all four Services and the US Coast Guard participated in this joint test and evaluation program sponsored by OSD. The purpose of the Throughput Phase of JLOTS II was to evaluate the Services' capabilities to discharge dry cargo and bulk POL from commercial and Navy ships at offshore anchorages with several cargo handling systems that had been developed over the previous 10 years. Military hardware, doctrine, and unit performance were evaluated. The levels of performance of individual sub-systems varied considerably. Some equipment performed better than expected, while others failed completely. As a result of JLOTS II, the Services and other DoD staff elements could now apply the lessons learned to eliminate unsatisfactory equipment, refine or modify marginal systems, and proceed with the procurement of effective systems.

Major conclusions of the JLOTS II exercises included the following:

(a). The demonstrated capability of the Services to discharge containerized and palletized cargo on a sustained basis is presently limited to Sea State 2 conditions and below (where wave heights do not exceed 3.5 ft). Although limited cargo movement could continue in Sea State 3, a number of improvements to equipment and refinement of procedures are required in order to achieve greater cargo throughput rates.

(b). The demonstrated capability to transfer bulk POL products from ship to shore was extremely limited or almost nonexistent. Accordingly, the Army and Navy must vigorously pursue the development of POL systems capable of greater stand-off distances and adequate mooring systems for commercial tankers that can reliably transfer bulk POL products to the shore.

(c). Concurrent with equipment improvement, the Services should update over-the-shore doctrine and operational procedures which recognize new sealift assets such as the Maritime Prepositioning Ships (MPS), the Fast Sealift Ships (FSS), and the Tactical Auxiliary Crane Ships (TACS).

(d). The JCS and UC should insure that LOTS capabilities of Active and Reserve forces are further enhanced through periodic and realistic Joint LOTS exercises.

USAEWES Technology Transfer Workshop

In December 1987, USAEWES convened a Joint Logistics Over the Shore (JLOTS) Technology Transfer Workshop for the purpose of bringing together the various DOD R&D agencies and the field operating units involved with JLOTS activities. Thirty-three different military organizations from Departments of Army and Navy, and Joint Command, attended and/or participated. Major representation was afforded by US Transportation Command, US Central Command, USA Logistics Center, USA Engineer School, USA Materiel Command, Naval Facilities Engineering Command, Naval Sea Systems Command, Amphibious Construction Battalions One, David Taylor Research Center, and Naval Civil Engineering Laboratory. Definition and prioritization of the technology base problems in the area of LOTS were discussed, and information exchange within the Military Research, Development, Test and Evaluation (RDT&E) community was conducted regarding present capabilities and on-going efforts. A 2-day round-table workshop involved key field representatives, mid-level program managers from various DOD agencies, and selected government contractors. Field elements provided a definition of their respective problem areas, and developed a framework for future JLOTS and amphibious operations RDT&E efforts. The research program presently being conducted by CERC was developed as a result of problem prioritization by this highly successful technology transfer workshop.

JLOTS III Feasibility Study

An OSD Joint Test and Evaluation (JT&E) Senior Advisory Council (SAC) met 15 August 1989 and set priorities for FY90 funding of feasibility studies of new projects. The first priority recommendation of the SAC was JLOTS III, due to broad support of CINCs, Services, and the JCS. The OSD-funded JLOTS III feasibility study was conducted during the first half of FY90 by the federally-funded research and development center, David Taylor Research Center of Annapolis, Maryland, using contractor support and with US Transportation Command (USTRANSCOM) oversight. JLOTS III has received strong support from the Army Deputy Chief of Staff, Logistics (DCSLOG), Navy OP-04, Joint Staff/J4, and ASD, Production and Logistics (PL). As

recognized by the OSD JT&E SAC, accomplishment of JLOTS III is the only way to thoroughly test the interoperability of systems obtained since JLOTS II, and determine whether planned tonnages can be delivered over the beach in quantities needed to sustain joint military operations. CERC is an active participant in the JLOTS III JT&E Feasibility Study, and will participate in the JLOTS III exercises scheduled for FY91 through FY94 by evaluating the effectiveness and performance of USACE-developed techniques for enhancing a LOTS operation.

USAEWES CERC LOTS Research and Development

Three work units of the 7-unit work package for sustainment engineering enhancement are presently on-going, with the remaining efforts to be initiated in FY91 and FY94. This will allow an evaluation of the effectiveness of the resultant RDT&E products during JLOTS III.

Sea State Data Base for LOTS

LOTS operations are critically dependant upon the environmental conditions at the site, including waves, winds, currents, tides, sea-bottom type, and the presence of bars or reefs. The objective of Sea State Data Base for LOTS is to provide the Army with a climatological sea state required for LOTS operations decisions and equipment design. This work is presently on-going. One critical geographic region is being addressed, and a methodology applicable for detailed site analysis is being established. Sea state can vary with time and over short spatial distances within the LOTS operational site. This work will provide a wave climate for one region to specify statistical information including percentage of time LOTS can be expected to operate, probability of site destruction by weather related events, and mix of equipment to be used in the operations.

Numerical models for predicting sea state will be used to develop a wave data set from historical weather observations in one area of possible LOTS application, in order to define the regional wave climate. A numerical model capable of defining in detail the nearshore operations area for LOTS will then be used to refine the wave data for two potential LOTS sites. The accuracy of the model will be demonstrated by comparisons to field data gathered in other areas, and the capability of the refined computations will be established by predictions during JLOTS III operations based on measured input and multiple observation sites within the operational area. The Persian Gulf and Gulf of Oman areas has been selected as the study site, in consultation with US Third Army and US Central Command. Data availability and quality evaluation is being conducted with cooperation from the US Air Force Environmental Technical Applications Center and the Naval Fleet Numerical Oceanographic Center. Relevant meteorological data has been acquired and is being processed.

The technique used for wind and wave condition investigation is hindcasting. This is the reconstruction of a long time-series of past conditions in an area. Hindcasting is almost always necessary for studies such as this because direct, long-term measurements of winds and waves

over a large region do not exist. The data that do exist (meteorological stations, ship observations, etc.) are used as input for numerical models of winds and waves. The principal activity to this point has been the preparation of hindcast wind fields for the Persian Gulf region. The hindcast consisted of converting meteorological data, which are point data for scattered locations, into a grid field of surface winds over the body of water for use in later wave hindcasting phases of the project. The procedure used is as follows.

(a). The time history of sea-level pressure and surface air temperature at each station is screened for bad data values. Identification of bad values is determined by impossible values of variables, and by inconsistency with nearby stations, and previous and subsequent values for the same station. Programs developed by project staff flag suspicious values which are then examined in detail.

(b). Short gaps (less than 12 hours) in the records for each station are filled by linear interpolation. This provides more station values for the spatial interpolation in the next task.

(c). The station values for isolated points are interpolated spatially to fill a grid, since the wind and wave hindcast models operate on grids covering the study area. The grid being used is a 50 x 100 point square grid with a step size of 16 km. A step size this small was needed to resolve the shape of the Persian Gulf. The spatial interpolation programs, which were developed during this research effort, produce a smooth surface (minimized mean square curvature at every point) which passes through each data point. No previously developed spatial interpolation scheme was found to be acceptable for producing data fields for use by the wind hindcast model, as all left artifacts in the fields (such as excessive local surface curvature) which would cause problems in the wind hindcast.

(d). The grids of surface pressure and air temperature, along with sea surface water temperature, are used as input for a wind hindcast numerical model. The model determines the geostrophic wind from the pressure field, corrects it for the gradient wind (acceleration due to curvature of the flow), and uses a boundary layer model incorporating both the planetary boundary layer and the turbulent frictional boundary layer to determine the surface (19.5 meter) wind for use by the wave hindcast model. This model was developed during an earlier study at CERC, and has been upgraded for the LOTS effort. The grid used in this phase is a 37 x 77 subgrid of the interpolation grid. The wind and wave grids need only cover the water area, while the interpolation grid covers all meteorological stations, some of which are inland.

The wind hindcasts developed to this point will be analyzed to provide wind statistics. The wind fields will be used in a wave hindcast for the entire Persian Gulf region using a shallow water wave hindcast model developed previously at CERC. This model will provide wave spectra and wave parameters over the regional grid. Sites for more detailed study of the nearshore region will be selected in consultation with US Third Army

and USACE. CERC-developed wave transformation models will be used to determine the effects of the local topography. The results of this study will be presented in a statistically analyzed format suitable for planning and operational use by US military elements responsible for LOTS conduct.

Water Levels and Currents

The objective of this work is to provide the Army with an operational numerical model for predicting currents and water levels for a geographic region of potential LOTS operation. This work is presently on-going. Predictions will be on a scale commensurate with specific LOTS sites. The model will be field validated at one potential site within the selected region. It also will be demonstrated at the selected site for the JLOTS III field exercises. Knowledge of currents and water levels expected in LOTS contingency areas will allow better coordination and planning of ship, pipeline, landing craft, and water purification operations, and also will be extremely helpful in predicting the movement of floating mines.

A global numerical grid has been established for a large area (Persian Gulf and the Gulf of Oman). The model is being calibrated and verified for currents and water levels driven externally by astronomical tide and wind. A site-specific numerical grid is being developed to cover a contingency area involving LOTS operations. Detailed site-specific tidal and wind-driven currents and water levels will be determined using boundary conditions provided by the global grid. The model also will be applied during JLOTS III field exercises.

Primary external wind forcing for the model will be provided by the wind fields developed in Sea State Data Base for LOTS. External tide forcing along the seaward edge of the Gulf of Oman will be provided by the global tide model developed at Naval Surface Weapons Center. The model has already been acquired by CERC. Data from these all available sources will be used in validating these external forcings, and verifying model predictions at interior grid points. Local wind information will be helpful in site-specific modeling in light of the strong sea breezes which characterize some seasons. Sea breezes can be inconsequential on the scale of a grid covering the Persian Gulf, yet dominate within 10 to 20 miles of the coast, the critical LOTS area.

Model. Selection of a model for water levels and currents is the next critical step. Important features of the Persian Gulf include strong stratification in some seasons, heavy loss to evaporation, temperature, and salinity differences likely to drive currents. The model selected should be capable of treating these features as needed. The model also must be capable of combining a coarse grid (Gulf-scale) with a fine grid (site-specific scale) in an effective way.

Two models are presently being considered as candidates. The model CH3D, developed at CERC, is a three-dimensional finite difference model capable of treating vertical structure in the water column. It does not presently contain some of the important features needed (e.g., evaporation), but these can be added forthwith. The model can be

configured to use information from a coarse grid to drive the seaward edge of a fine grid.

The other candidate model is a finite element model being developed jointly by Texas A&M University and CERC under the Dredging Research Program. The model has as powerful capability for using triangular grid elements of widely varying size, which allows large elements in offshore waters, and with increasingly smaller grid elements in areas where water depth decreases. The triangular elements will fit an irregular shoreline well. This model will allow regional and site-specific results to be obtained with one grid in a single computer simulation, with savings in both time and costs. The present model does not include vertical structure; however the potential for adding this feature is being investigated. Developmental work is continuing on both candidate models. Both will be evaluated and tested for LOTS application. A selection will be made and a final regional grid developed in late FY90.

Regional Modeling. The selected model will be extensively tested on the Persian Gulf regional grid. Model results at interior grid points near data sources will be used to test wind forcing. Events dominated by winds will be used to test wind forcing. Events dominated by astronomical tides will be used to test tidal forcing. General and seasonal trends evidenced in the model will be compared with known patterns. When the model has been sufficiently verified, it will be used to simulate the time period 1973-1986, also covered by Sea State Data Base for LOTS. Water level and current time histories will be archived at relevant points.

Site-Specific Modeling. A potential LOTS site in the Persian Gulf will be selected. A fine-scale numerical grid will be developed, tested, and verified as needed to cover the nearshore region. Special computational routines will be added to identify and treat the surf zone. Both regional and site-specific information will be summarized in an optimum format for LOTS planning purposes. A customized data display will be developed to facilitate and enhance CINC planning processes.

Simulation Planning Model for LOTS

The objective of this research effort is to provide the Army a numerical logistics simulator for evaluation of LOTS operational transfer of military equipment and material under adverse conditions in remote areas of operation. This work is presently on-going. Analyses are being conducted which include environmental requirements, vessel forces, mooring considerations, wave climate and attenuation, geographical, and material requirements. Results from these analyses will be incorporated into a numerical logistics simulator which will model the entire LOTS operation. The model will then be tests and validated in coordination with JLOTS III exercises. Final products will be software and user guides.

Coastal Environmental Evaluation

The objective is to provide the Army a computer-based method to serve as a tool for prioritization and selection of LOTS sites, and design of

support equipment. This work will begin in FY91. To fully consider
locations where LOTS sites are feasible, coastal conditions as well as sea
conditions must be evaluated. A successful site will allow effective
movement of military cargo and require little maintenance for the facility
to remain operational. Evaluation of wave conditions, sedimentation
patterns, and the geomorphologic settings must be combined to thoroughly
characterize a coastal environment. Depending upon the type or category
of coastal environment, different LOTS associated equipment will be
required to construct, operate, and maintain a facility effectively.
Methodology will be developed to inventory and categorize potential LOTS
coastal environments from information available for specific sites or
entire regions. Coastal areas will be categorized based upon bathymetry,
wave, water level, sediment characteristics, and other physical
parameters. This will aid in the selection of possible LOTS sites, and
identification of any special equipment necessary for a particular
category of coastal environment. Data from other work units and existing
data bases will be incorporated with reconnaissance data to provide a
worldwide classification criteria of LOTS sites.

Real-Time Forecasting of Sea States

The objective is to provide the Army a field deployable microcomputer
sea state prediction system to forecast sea conditions vital to operations
in offshore and nearshore areas and ports around the world. This work
will begin in FY93. The system will be based on proven, operational Army
sea state prediction models. It will be validated at a field site. It
also will be tested during JLOTS III field exercises if development is
sufficiently adequate. The sea state prediction system will allow
commanders to anticipate LOTS interruptions, and how long they will last.

A suitable source of meteorological forecasts will be identified and
interfaced with an available sea state numerical model. Useful forecasts
for 24-72 hours in advance can be expected. The linked meteorological/sea
state model will be verified using extensive intercomparison of
predictions to measurements. The system will be tested during JLOTS III
exercises if at a suitable state of development. The model will be
adapted for use on microcomputer equipment available to field commanders.
User-friendly interfaces and output products customized for sponsor needs
will be developed. Training procedures for the system will be developed.

Rapidly Erectable Breakwaters (REB) for LOTS and Ports

The objective of this research unit is to develop and evaluate
effective design concepts for rapidly-deployable breakwater systems for
military application for LOTS operations protection. This work will begin
in FY92. Such operations rely on temporary facilities across the surf
zone which will be enhanced by effective mobile breakwaters to attenuate
waves from Sea State 3 to an operable Sea State 2. Analytical and
numerical model studies will optimize mobile breakwater geometry and
design. The optimum concept based on military requirements, logistical
considerations, time for deployment, and cost will be tested at reduced

scale in the physical test model laboratory to verify conceptual design parameters.

A review of existing mobile (floating) breakwater technology was conducted previously under another aspect of floating breakwater research activities. Environmental loadings of current, wind, and wave forces for design criteria were determined. A Hybrid Green Function Model was determined superior for flexibility in handling arbitrary structure shapes and water depths. The numerical model was made user-friendly, and was used to initiate optimum breakwater configuration designs. The new generation of pontoon causeway sections (5' x 8' x 40') is considered a potentially viable structural element for fabrication into prototype mobile breakwater units. Other flexible, collapsible membrane fabric elements also will be evaluated because of the ability to transported large quantities in relatively small container volumes.

Field Validation of REB

The validation of newly developed equipment and facilities under actual field application determines the usefulness to the proponent agency. This work will be initiated in FY94. The objective of this research unit is to evaluate the feasibility and reliability of utilizing rapidly-deployable breakwater systems to attenuate open ocean wave climates for LOTS operations from presently inoperable Sea State 3 to a documented level of Sea State 2. A typically representative section of optimum breakwater design will be mobilized by container-capable logistical equipment, transported to a field evaluation site, installed during on-site engineer training, monitored for wave attenuation effectiveness while on station, and retrieved for later utilization at other LOTS sites. Correlation of prototype wave attenuation, member stresses, anchor forces, and connector fatigue with laboratory reduced scale tests will be determined. Appropriate modifications of structural elements based on prototype evaluation will be effected.

Prototype module units developed by Rapidly Erectable Breakwaters (REB) for Ports and LOTS will be acquired to fabricate a demonstration mobile breakwater section. Necessary logistic support to install, monitor, and retrieve the prototype section will be acquired. Training procedures and manuals for engineer field unit installation will be developed. Structure field demonstrations will be performed with on-site engineer training. Installation and training procedures will be refined based on prototype installation experience, and doctrinal manuals will be updated as necessary.

Conclusions

As previously expressed by Rear Admiral Thomas Hayward, former Chief of Naval Operations, "...Sealift is the critical element of strategic mobility. Without adequate and reliable sealift, none of our military plans are executable..." The USACE is fully cognizant of the ramifications associated with Combat Service Support and sustainment of deployed ground forces in the field. USACE has committed significant

funding to address demonstrated deficiencies which limit present LOTS capabilities for throughput of material necessary to maintain field units. Results from these RDT&E activities will greatly enhance Service capacity for sustainment by significantly reducing LOTS interruptions due to inclement environmental factors.

BIBLIOGRAPHY

Atturio, M. et al, _Concepts for Improving Logistics Capabilities Over the Shore_, N-1738, Naval Civil Engineering Laboratory, Port Hueneme, CA, November 1985.

Beakey, D.J., _Logistics Over The Shore: Do We Need It?_, Monograph Series 82-6, National Defense University Press, Fort McNair, Washington, D.C., 1982.

Hales, L.Z., _Floating Breakwaters: State-of-the-Art Literature Review_, TR 81-1, USA Coastal Engineering Research Center, Fort Belvoir, VA, October 1981.

Marino, J.N., _Floating Breakwaters: Various Designs and Possible Military Uses_, Requirement Report, USA Engineer Advanced Course, Fort Belvoir, VA, 17 February 1982.

Privratsky, K.L., _British Combat Service Support During the Falkland Islands War: Considerations for Providing Operational Sustainment to Remote Areas_, USA Command and General Staff College, Fort Leavenworth, KS, 1 April 1986.

Russell, J.A., "Deployment: Will TRANSCOM Make a Difference?", _Military Logistics Forum_, June, 1987, p. 38-45.

Wullenjohn, C., "Services Cooperate in LOTS Operation," _Army Logistician_, USA Logistics Management Center, Fort Lee, VA, p. 14-16.

_____, _Joint Logistics Over-the-Shore II Test and Evaluation: Deployment Test Report_, DOD Joint Test Directorate, Little Creek Naval Amphibious Base, Norfolk, VA, August 1985.

_____, _Joint Logistics Over-the-Shore II Test and Evaluation: Roll-On/Roll-Off Ship Operations_, DOD Joint Test Directorate, Little Creek Naval Amphibious Base, Norfolk, VA, March 1984.

_____, Joint Logistics Over-the-Shore II Test and Evaluation: Photographic Summary Report, DOD Joint Test Directorate, Little Creek Naval Amphibious Base, Norfolk, VA, July 1985.

_____, Joint Logistics-Over-the-Shore II Test and Evaluation: Throughput Phase, DOD Joint Test Directorate, Little Creek Naval Amphibious Base, Norfolk, VA, August, 1984.

_____, Joint Logistics Over-the-Shore II Test and Evaluation: Analysis and Evaluation Report, DOD Joint Test Directorate, Little Creek naval Amphibious Base, Norfolk, VA, August 1985.

_____, Joint Strategic Sealift Offshore Discharge Operations, Coordinating Draft, FM 55-64, US Army, Volumes 1&2, 15 January 1987.

_____, Joint Logistics Over The Shore III Draft Feasibility Study, Working Copy, US TRANSCOM, Scott AFB, IL, 12 February 1990.

Laser Photochemical Studies of Intermediates in
Energetic Materials Combustion

Michael J. McQuaid,[a] Dr.
*Rosario C. Sausa, Dr.
Andrzej W. Miziolek, Dr.
Clifton N. Merrow,[b] Dr.
Ballistic Research Laboratory, SLCBR-IB-I,
Aberdeen Proving Ground, MD   21005-5066
a. NAS/NRC Postdoctoral Research Associate
b. Chemical Research and Development and Engineering
Center, SMCCR-DDT, Aberdeen Proving Ground, MD 21010-5423

INTRODUCTION: Cyclic nitramines such as HMX and RDX are energetic materials used in propellants and explosives. One of the keys to understanding the ignition and combustion of these compounds is a characterization of the pathways involved in their unimolecular decomposition. This information is difficult to obtain since the utility of energetic materials is based on their propensity to undergo extremely rapid, exothermic reactions. Moreover, functionally useful nitramines may decompose via a number of possible pathways, and secondary bimolecular reactions involving product radicals may obscure the identity of the initial reaction products.

Because of these difficulties dimethylnitramine (DMNA), $(CH_3)_2NNO_2$, has been studied as a simple analog expected to exhibit some of the reactions important in cyclic nitramine decomposition. DMNA decomposition has been investigated in thermal pyrolysis, photodissociation and theoretical studies. UV laser photochemistry has proven to be a valuable tool for chemical analysis,[1] for decomposition studies of chemical warfare agent simulants,[2] as well as for the detection of energetic materials molecular fragments.[3] One of the early investigations of DMNA decomposition was a study reported by Fluornoy on the thermal pyrolysis of gaseous DMNA in a static bulb.[4] This researcher found that approximately 80% of the pyrolyzed DMNA was converted to dimethylnitrosamine, $(CH_3)_2NNO$. Fluornoy suggested that the rate determining step for DMNA dissociation was a scission of the $(CH_3)_2 N-NO_2$ bond,

$$(CH_3)_2N-NO_2 \rightarrow (CH_3)_2N + NO_2 \qquad (1)$$

although this was not directly demonstrated. More recent studies have corroborated the importance of this reaction step. Lin and coworkers came to this conclusion based on the results of low temperature static bulb and high temperature single pulse shock tube experiments.[5] Wodtke and Lee have made preliminary reports of infrared multiphoton decomposition of DMNA in a molecular beam leading to the conclusion that reaction (1) is the exclusive (>95%) initial step.[6]

Golden and coworkers have studied the decomposition of DMNA via (1) GC analyzed laser pyrolysis experiments in a flow cell,[7],[8] and (2) mass spectrometric sampling of DMNA pyrolyzed in a molecular beam.[9] These researchers have concluded that a nitro-nitrite rearrangement,

$$(CH_3)_2N-NO_2 \rightarrow (CH_3)_2NO-NO \tag{2a}$$

followed by rapid scission of the $(CH_3)_2NO-NO$ bond,

$$(CH_3)_2NO-NO \rightarrow (CH_3)_2NO + NO \tag{2b}$$

effectively competes with reaction (1).

Another initial step which has received consideration is HONO elimination,

$$(CH_3)_2NNO_2 \rightarrow (CH_3)N(CH_2) + HONO \tag{3}$$

Shaw and Walker have postulated that HONO elimination contributes no more than 10% to the rate of DMNA decomposition.[10] Similarly, Sumpter and Thompson find that HONO elimination is not important for a random distribution of energy, but that the rate of this reaction is significantly increased by overtone excitation of one of the CH stretches.[11]

In order to shed new light on the physical and chemical processes occurring in the thermal decomposition of DMNA, it is worthwhile studying the decomposition of DMNA following its photolysis with UV radiation. Mialocq and Stephenson demonstrated the value of this approach in their (single frequency) picosecond laser study of the collisionless photodissociation of DMNA.[12] These researchers found that reactions,

$$(CH_3)NNO_2 \xrightarrow[226 \text{ nm}]{h\nu} (CH_3)_2N + NO_2 \tag{1a}$$

$$\xrightarrow[226 \text{ nm}]{h\nu} (CH_3)_2N + NO_2^* \tag{1b}$$
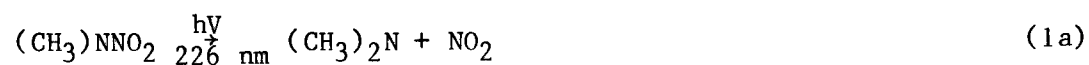
were monophotonic, collision-free and occurred in 6 ps or less. These results showed that UV photolysis of DMNA yielded a pathway considered important in thermal pyrolysis, and emission, as well as LIF spectroscopy allowed the dissociation products to be followed in "real time".

In this paper, the results of photolyzing DMNA with 248 nm laser radiation are presented. Emission features in the range 255-855 nm resulting directly from DMNA photolysis have been recorded and assigned. Tunable laser induced excitation was used to probe for ground state OH, NO, and NO_2 formed by the photolysis. The results demonstrate the utility of this approach for characterizing DMNA decomposition pathways. The knowledge gained will also be valuable for the extension of these techniques to more complex nitramine systems.

EXPERIMENTAL METHODS: The experimental apparatus used for this study is depicted in Figure 1. An excimer laser (Lumonics, HyperEX 440), usually operated at 10 Hz with KrF (248 nm), was used to irradiate a "slow flow" of pure DMNA vapor. Slow flows were utilized to avoid the buildup of reaction products and/or depletion of reactants. As an additional precaution against reaction product build-up, the system was flushed with Ar and then completely evacuated between experimental runs. The DMNA vapor was introduced into the vacuum chamber via a valve on an evacuated flask containing the degassed, crystalline compound. The valve was usually full open and the flow of effluent to the vacuum pump was throttled to obtain the desired sample pressure (10-100 mTorr). A liquid nitrogen trap was installed between the vacuum pump and the effluent throttle valve to prevent pump oil from backstreaming into the flow cell. The DMNA sample was synthesized at ARDEC, and mass and IR absorption spectra were taken to verify its purity.

The photolysis beam entered and exited the vacuum chamber via MgF_2 windows. The unfocussed beam was collimated with irises to a 1 cm diameter. It was found that photolysis energies ≤5 mJ/pulse were necessary to avoid the corrupting influence of multiphoton dissociation processes. Fluorescence resulting directly from the photolysis event was collected at 90° to the beam by a 3" fused silica lens and exited the vacuum chamber via a MgF_2 viewing port. The emission was then focused on the entrance slit of a monochromator (Spex, 500M) equipped with an EMI 9558QB photomultiplier tube (PMT). The spectral response of this arrangement was calibrated using a standard lamp (Optronic Laboratories) and a mercury lamp.

The output signal from the PMT was directed to a gated integrator (Stanford Research Systems) and displayed on a digital oscilloscope (Hewlett-Packard). The temporal delay between the photolysis and probe pulses, and radiative lifetimes of the fluorescing products, was

determined from the oscilloscope display. Spectra were recorded in digital form on a computer via a commercially obtained computer interface/software package (Stanford Research Systems).

For the laser induced fluorescence experiments, the probe beam was provided by an XeCl excimer-pumped dye laser (Lumonics HyperEX 440/HyperDYE 300) which was frequency doubled to obtain $\lambda <$ 310 nm. The probe beam was collinear and counterpropagating with respect to the photolysis beam. The delay between the arrival of pump and probe pulses was controlled by a digital delay generator (Stanford Research Systems). Fluorescence was collected as above and the monochromator used as bandpass filter to block laser scatter and extraneous fluorescence. Alternately, the PMT could be placed adjacent to the viewing port with appropriate "glass" bandpass filters. The following detection schemes were used for probes of OH, NO, and NO_2:

OH: The OH $A^2\Sigma^+ - X^2\Pi_i$ (1,0) and (0,0) electronic transitions are excited in the ranges 281-285 nm and 306-310 nm, respectively. When exciting the (1,0) transition, emission corresponding to the (1,1) transition was monitored using a band pass filter (Corion) centered at 313 nm. When exciting the (0,0) transition, emission corresponding to the (0,1) transition near 347 nm was monitored using a WG-335, UG-11 (Schott) and 350 nm (Optics Technology) bandpass filter combination.

NO: The NO $A^2\Sigma^+ - X^2\Pi$ (0,0) electronic transition is excited near 226 nm. In this case, emission associated with the (0,1) transition was monitored near 236 nm using the monochromator as a bandpass.

NO_2: The NO_2 A-X transition was excited in the ranges 530-550 nm and 440-450 nm. Fluorescence in the range $\lambda \geqslant$ 600 nm was monitored using various combinations of long pass and bandpass filters.

The photon dependencies of the pathways leading to the various reaction products were measured by varying the fluence of the photolysis laser using calibrated dielectric filters (Acton Research).

RESULTS: The absorption coefficient for DMNA at 248 nm is 6.7 ± 0.7 x 10^{18} $cm^2 molecule^{-1}$.[10] The emission spectrum observed following the photolysis of DMNA at 248 nm is shown in Figure 2. This spectrum is a composite of scans taken for $\lambda <$ 400 nm and $\lambda >$ 400 nm. The relatively long radiative lifetime of the features ascribed to NO_2 $\tilde{A}$ ($^2B_2$) $\rightarrow$ $\tilde{X}$ ($^2A_1$) emission, compared to other emission features, necessitated that higher signal gains be used for recording this portion of the spectrum. Because of the differing radiative lifetimes, comparison of the relative intensities is not straightforward. The emission feature centered at 308 nm is due to OH $A^2\Sigma^+ - X^2\Pi_i$ (0,0) emission. This assignment is based upon a computer simulation of the

feature (see Figure 3) and the observed radiative lifetime. The features centered at 288 nm and 298 nm have not been assigned with certainty. These features have a radiative lifetime of ~100 ns at 20 mT of DMNA. Further investigations attempting to characterize the carrier of these features are being pursued.

A log-log plot of the integrated emission intensity as a function of photolysis laser energy for the features ascribed to OH and NO_2 are presented in Figure 4. For OH, the plot yields a slope of 2.1 ± 0.2 implying that formation of the A state requires two photons. For NO_2, the plot yields a slope of 1.0 ± 0.1, implying the NO_2 $\tilde{A}$ state is formed in a one photon process. The latter result is similar to the result obtained by Mialocq and Stephenson for DMNA photolysis at 266 nm. However, we were unable to detect ground state NO_2 via laser induced excitation at 530 nm. Fluorescence from the NO_2 $\tilde{A}$ state following 448 nm excitation has been observed, but these results are preliminary.

A portion of the laser induced excitation spectrum of ground state OH obtained following DMNA photolysis is shown in Figure 5. The delay between the arrival of the photolysis and probe pulses was 50 ns. A log-log plot of the integrated intensity of the LIF signal for excitation of the $R_1(6)$ bands as a function of photolysis laser energy yields a slope of 1.0 (see Figure 4). These results, which were obtained under conditions where collisions should not be important, indicate that ground state OH is produced from a primary photodissociation step in a single photon process. Analysis to correlate the excitation spectrum with the nascent OH rotational distribution is currently in progress. No attempt was made to look for OH vibrational excitation in this work.

LIF excitation spectra were obtained for ground state NO. However, this spectra could be obtained in the absence of a photolysis event. Attempts to isolate signal due solely to photolysis products have been unsuccessful. DMNA, NO_2, and HONO all have the potential to absorb 226 nm radiation and dissociate to form NO. Attempts to circumvent this problem by using a 2 photon excitation scheme (2 x 452 nm) have been similarly unsuccessful. Since it was considered unlikely that the multiplicity of pathways leading to NO formation could be sorted out based on LIF detection, further efforts in this direction have been suspended.

DISCUSSION: The inability to excite ground state NO_2 at 530 nm following DMNA photolysis with 248 nm radiation raises intriguing possibilities regarding mechanistic differences for DMNA photolysis with 248 nm (5.0 eV) and 266 nm (4.7 eV) radiation. It may be that a significantly larger fraction of NO_2 is formed in the A state with the higher energy photons. Further experiments will be performed in this

laboratory to establish the nature of the differences in the photolysis at these two wavelengths.

The observation of OH $X^2\Pi_i$ via a one photon process and OH $A^2\Sigma^+$ via a two photon process would be consistent with the formation of an HONO* intermediate. The activation energy for the HONO elimination channel has been estimated to be approximately 40 kcal/mole (1.7 eV) and the reaction slightly exothermic.[10,14] The absorption of a 5.0 eV photon is sufficient to overcome the activation energy and break the 50 kcal/mole (2.09 eV) HO-NO bond.[15] HONO has also been shown to absorb at 248 nm.[16] The formation of OH $A^2\Sigma^+$, v=0 state in the two step process,

$$(CH_3)_2NNO_2 \xrightarrow[248\ nm]{h\nu} (CH_3)N(CH_2) + HONO \tag{4a}$$

$$HONO \xrightarrow[248\ nm]{h\nu} OH(A) + NO(X) \tag{4b}$$
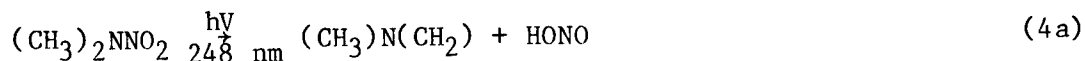
is energetically feasible for HONO formed with $\geqslant 1$ eV of internal excitation. It should be noted that ground state OH may absorb 248 nm. However, excitation at this wavelength would lead to the short-lived, predissociative v'=3 level of the OH A state. This state would yield a significantly different spectrum than shown in Figure 2.

Two reaction mechanisms have been proposed for the formation of ground state NO, i.e., reactions (2) and (3). The inability to discern nascent NO formation via LIF prevented any correlation between its formation and OH formation; which in turn, may have provided insight into the relative importance of the competing channels. Alternative methods for detecting NO in this environment are currently being investigated.

CONCLUSION: The UV photolysis of DMNA initiates its unimolecular decomposition in a low pressure, collision-free environment. Using emission and LIF spectroscopy, the products of various decomposition pathways can be observed on a time scale precluding the involvement of products from secondary bimolecular reactions. Thus, the ambiguity inherent in end product analysis is avoided. Pathways leading to the formation of OH $A^2\Sigma^+$, OH $X^2\Pi_i$, and NO_2 $\tilde{A}$ (^2B) have been investigated in this study. The formation of OH in the photolysis of DMNA is consistent with a HONO elimination channel. The results indicate that the pathways initiated by the UV photolysis of DMNA are analogous to the pathways expected in the thermal decomposition of this molecule. As such, UV photolytic studies provide a knowledge base for thermal decomposition investigations.

REFERENCES:

1. R.C. Sausa, A.J. Alfano, and A.W. Miziolek, "Efficient ArF Laser Production and Detection of Carbon Atoms from Simple Hydrocarbons", Appl. Opt., Vol. 26, p. 3588, 1987.

2. R.C. Sausa, A.W. Miziolek, and S.R. Long, "State Distribution, Quenching and Reaction of PO Generated in Excimer Laser Photofragmentation of Dimethyl Methyl Phosphonate", J. Phys. Chem., Vol. 90, p. 3994, 1986.

3. P.J. Dagdigian, W.R. Anderson, R.C. Sausa, and A.W. Miziolek, "The Photodissociation of Formaldoxime and Its Methylated Homologs: Search for H_2CN Fluorescence", J. Phys. Chem., Vol. 93, p. 6059, 1989.

4. J.M. Fluornoy, "Thermal Decomposition of Gaseous Dimethylnitramine", Journal of Chemical Physics, Vol. 36, p. 1106, 1962.

5. S.A. Lloyd, M.E. Umstead, and M.C. Lin, "Kinetics and Mechanism of the Thermal Decomposition of Dimethylnitramine at Low Temperatures", Journal of Energetic Materials, Vol. 3, p. 187, 1985.

6. A.M. Wodtke, private communication (see References 4 and 5).

7. D.F. McMillen, S.E. Nigenda, A.C. Gonzalez, and D.M. Golden, "Pulsed Laser Pyrolysis of Dimethylnitramine and Dimethylnitrosamine", Spectrochimia Acta, Vol. 43A, p. 237, 1987.

8. S.E. Nigenda, D.F. McMillen, and D.M. Golden, "Thermal Decomposition of Dimethylnitramine and Dimethylnitrosamine by Pulsed Laser Pyrolysis", Journal of Physical Chemistry, Vol. 93, p. 1124, 1989.

9. P.H. Stewart, J.B. Jeffries, J.M. Zellweger, D.F. McMillen, and D.M. Golden, "Molecular Beam Sampled Laser Pyrolysis of Dimethylnitramine", Journal of Physical Chemistry, Vol. 93, p. 3557, 1989.

10. R. Shaw and F.E. Walker, Journal of Physical Chemistry, Vol. 81, p. 2572, 1977.

11. B.G. Sumpter and D.L. Thompson, "Unimolecular Reaction Dynamics of Dimethylnitramine", Journal of Chemical Physics, Vol. 88, p. 6889, 1988.

12. J-C. Mialocq and J.C. Stephenson "Picosecond Laser Study of the Collisionless Photodissociation of Dimethylnitramine at 266 nm", Chemical Physics Letters, Vol. 123, p. 390, 1986.

13. R.C. Sausa (unpublished results).

14. C.F. Melius and J.S. Binkley, "Thermochemistry of the Decomposition of Nitramines in the Gas Phase", Twenty-First Symposium on Combustion, pp. 1953-1963, The Combustion Institute, Pittsburgh, 1986.

15. JANAF Thermochemical Tables, Dow Chemical Co., 1960-1963.

16. R.A. Cox and R.G. Derwent, "The Ultra-Violet Absorption Spectrum of Gaseous Nitrous Acid", Journal of Photochemistry, Vol. 6, p. 23, 1976.
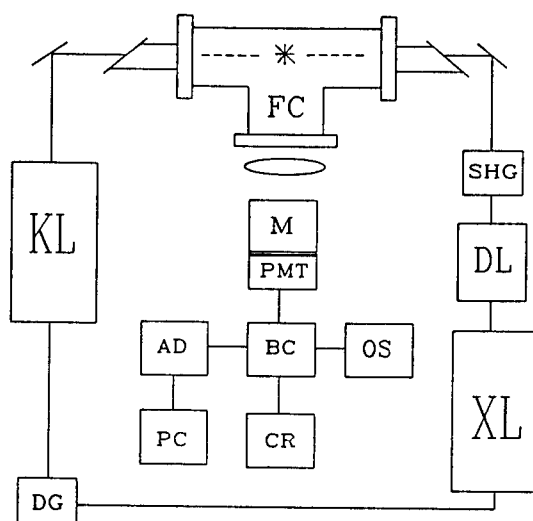
Fig. 1. Experimental apparatus for the emission and LIF study of DMNA photolysis products. Equipment includes: a KrF excimer laser (KL), a XeCl excimer laser (XL), a tunable dye laser (DL), a second harmonic generator (SHG) and a pulse delay generator (DG). Fluorescence in the flow cell (FC) is monitored through a monochomator or band pass filter (M) with a photomultiplier tube (PMT). Signal processing and recording is accomplished with a boxcar integrator (BC), A/D converter (AD), personal computer (PC), chart recorder (CR) and oscilloscope (OS).
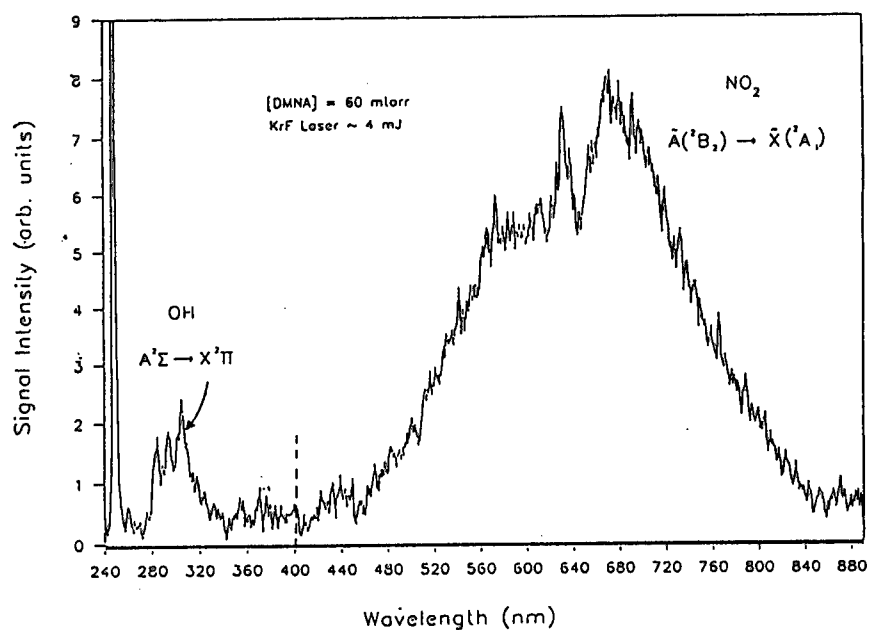
Fig. 2. Emission spectrum following DMNA photolysis using 248 nm laser radiation. The emission is monitored between 50 and 100 ns following the arrival of the photolysis pulse.
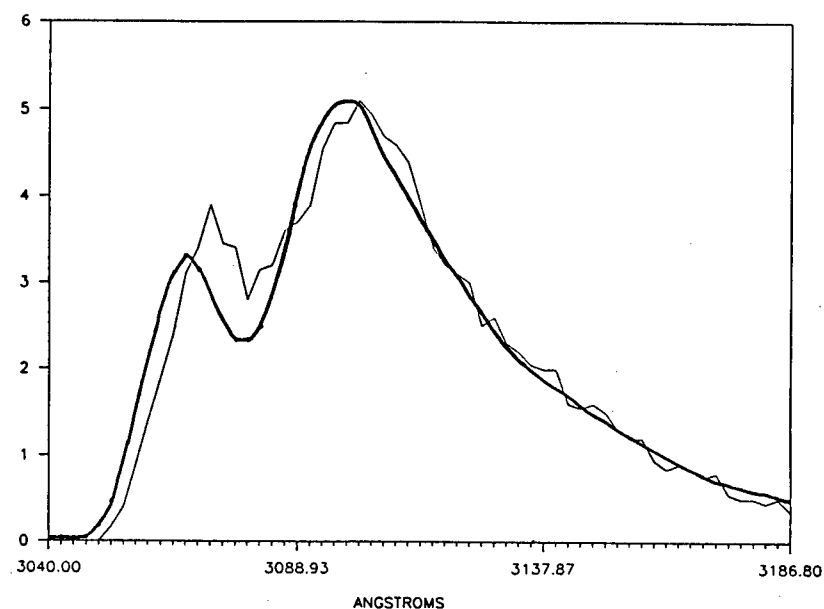
Fig. 3. Comparison of (1) the experimentally observed emission feature at 309 nm following DMNA photolysis (light line) and (2) a computer generated simulation of emission for the OH $A^2\Sigma^+$, v'=0 state (dark line). The simulation assumes a Boltzmann distribution of the OH A state rotational levels. The temperature corresponding to this distribution is 2500° K.
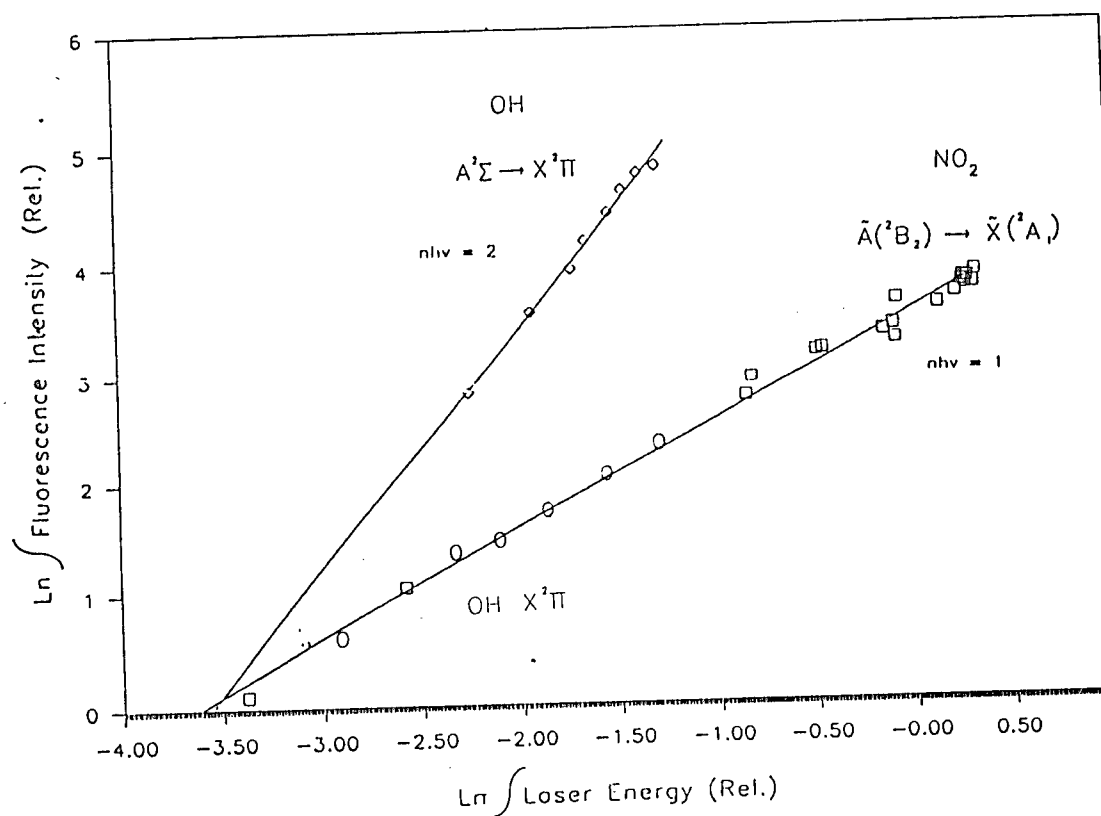
Fig. 4. Photon dependencies for the formation of the OH $A^2\Sigma^+$, OH $X^2\Pi_i$, and NO_2 A ^{2}B states. The fluorescence intensity of the OH $X^2\Pi_i$ state follows laser induced excitation of the OH $A^2\Sigma^+$ - $X^2\Pi_i$ (1,0) electronic transition.

OH Excitation Spectra
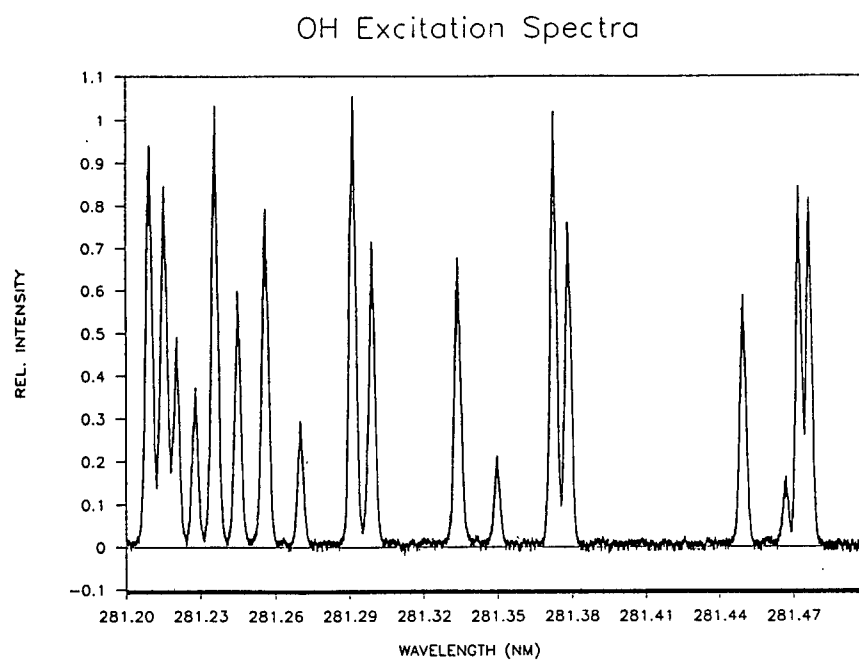
Fig. 5. Excitation spectrum near the OH A-X (1,0) band head
following photodissociation of DMNA with 248 nm radiation.

Tactical Estimate of the Situation: Past, Present, and Future (U)

Mr. Rex Michel
Army Research Institute Field Unit
P.O. Box 3407
Fort Leavenworth, Kansas 66027-0347

Introduction

The "estimate of the situation" is the label describing the doctrinally approved method for arriving at tactical decisions in the U.S. Army. It has a history of nearly a century of instruction in the major Army schools. There is scarcely an officer in the Army who has not been exposed to its precepts and practiced its implementation at sometime in his or her career.

Figure 1 illustrates the general procedures involved in the estimate.

1. Analyze the mission

2. Analyze the situation and develop possible courses of action

3. Analyze each course of action separately

4. Compare the courses of action

5. Decide which course of action to implement.

Figure 1. The basic 5-step estimate of the situation.

Much more detail than this, of course, is contained in the doctrinal descriptions of the process.[1] If followed, the final course of action selection is a product of pure deductive reasoning. Such a method is relatively easy to teach and to demonstrate, it can be made to account for virtually every nuance on the battlefield, and it leaves an "audit trail" by which the final decision can be justified. Its resulting instructional value and appeal to professionalism help to account for its long standing institutionalization in the Army.

A complete estimate process, however, is difficult to accomplish within the time constraints and complexity of modern warfare. Added to this is

mankind's propensity to generalize, assume, take shortcuts, and perseverate on the first idea that comes to mind. There needs to be, therefore, some matching of the estimate process to the realities of the modern battle-field. This paper looks at the historical development of the estimate process, its current status, and then offers an approach to adaptation to modern warfare constraints.

History of the Estimate

Attempting to understand your enemy, what he is likely to do, and how you might best defeat him have always been characteristics of good general-ship. But it was Napoleon who apparently first systematized the process. A written "estimate" from Napoleon's Leipzig Campaign in August, 1813 con-tains the analysis of five possible enemy courses of action and their re-lationship to his own course of action.[2]

The death of Frederick the Great and subsequent defeats made the Prussians realize how dependent they were upon the rare chance existence of true tactical genius; they became interested in developing a documented, systematic procedure "to develop by training a high average of ability in leadership." At the Prussian Kriegsacademie a system for teaching the art of command was evolved over about a half century. Called the "applicatory system", it included a systematic mental survey of the existing situation leading up to and expressed in the decision.[3]

The applicatory system was imported to the United States by Secretary of War, Emory Upton in 1875 who noted that it "taught good thinking habits to insure action from foresight rather than impulse."[4] When The School for Application of Infantry and Cavalry opened at Fort Leavenworth in 1881, there began a gradual adaptation of the applicatory system to the American psyche. The result was the publication of the first official description of the estimate process in the 1910 Army Field Service Regulations.[5] Although the description consisted of a single paragraph, it contained the essence of the estimate process used today.

Since 1910 the evolution of the estimate process has involved the gradual fleshing out of procedures and details for accomplishing the indi-vidual estimate steps. The concept and purpose of the estimate have re-mained intact. The admonition to consider all the pertinent facts, to gen-erate courses of action from this review and to compare them to arrive at a decision have remained stable since its inception some eighty years ago.

The doctrinal evolution of the estimate process since 1932 has been traced through nine sequential editions of the Army's doctrinal publication regarding staff procedures, Field Manual (FM) 101-5, Staff Organization and Operations.[6] Within the six editions of FM 101-5 between 1932 and 1968 there was a fairly consistent increase in the amount of detail given re-garding the estimate process. Both "how to" information and the types and amount of data to be used generally increased throughout this time period. Figure 2 illustrates the increase in the outline of the estimate between 1932 and 1968. The subsequent three editions, however, have added rela-tively little procedural information or additional data requirements. In

---- 1932 ----

1. MISSION
2. OPPOSING FORCES
   a. Enemy Forces
   b. Own Forces
   c. Relative combat strength
3. ENEMY SITUATION
   a. Plans open to the enemy
   b. Analysis of the enemy plans
   c. Enemy possible intentions
4. OWN SITUATION
   a. Plans open to you
   b. Analysis of plans open to you
5. DECISION

---- 1968 ----

1. MISSION
2. SITUATION AND COURSES OF ACTION
   a. Considerations affecting the possible COAs
      (1) Characteristics of the area of operations
         (a) Weather
            Military aspects and light data
            Effects on enemy and friendly operations
         (b) Terrain
            Effects on military aspects
            Effects of military aspects on Enemy/Friendly
            Key terrain features and avenues of approach
         (c) Other pertinent factors
      (2) Enemy situation
         (a) Dispositions
         (b) Composition
         (c) Strength
            Committed forces
            Reinforcements
            Air and NBC
            Other considerations
         (d) Recent and present significant activity
         (e) Peculiarities and weaknesses
      (3) Own situation
         (a) Dispositions
         (b) Composition
         (c) Strength
            Committed forces
            Reinforcements
            Air and NBC
            Other considerations
         (d) Recent and present significant activity
         (e) Peculiarities and weaknesses
      (4) Relative combat power
   b. Enemy capabilities
   c. Own courses of action
3. ANALYSIS OF OPPOSING COURSES OF ACTION
   a. List of enemy capabilities that will
      materially assist in choosing the best COA.
   b. Analysis of each COA verses each listed
      enemy capability.
4. COMPARISON OF OWN COURSES OF ACTION
5. DECISION

Figure 2. Comparison of estimate outlines -- 1932 versus 1968.

fact, they seem to retreat from previous imperatives to "consider all the facts that can affect accomplishment of the mission"[7] and the example estimates given, if any, are shorter and contain fewer facts than previous ones. It may be that the estimate process was considered adequately detailed by 1968. It may also be that the burgeoning amounts of information becoming available could quickly overwhelm a commander attempting to analyze "all the facts." A warning regarding data inundation was explicitly stated in the 1982 edition.[8]

Current Status of the Estimate

The estimate process remains firmly entrenched as the Army's doctrinally recognized method of selecting a tactical course of action. There has been no published revision to FM 101-5 since 1984, and this last edition is very similar to the one published in 1968 regarding the estimate of the situation. Confidence in the efficacy of the estimate process is expressed in such statements as "It is neither possible nor practical for the commander to reach a conclusion on the best course of action until all the information developed during analysis is available and the comparison (of courses of action) in paragraph 4 has been completed."[9]

There has been, however, considerable development of the estimate process going on at the Command and General Staff College (CGSC) at Fort Leavenworth since 1984. CGSC Student Text, 100-9, The Command Estimate, has undergone two revisions since then. ST 100-9 focuses on what it calls "the command estimate" which expands the estimate of the situation described in FM 101-5 by including the post-decision activities of developing and disseminating orders and supervising execution. It also includes commanders' intent and guidance which are missing from the FM 101-5 estimate. In effect, it combines the estimate of the situation with what FM 101-5 describes as the military decision making process. Yet ST 100-9 also claims to improve upon the estimate of the situation by reducing time requirements and avoiding a rigid structure. It does this by "highlight(ing) the information the commander and principal staff need to make timely accurate decisions" and "provid(ing) a means for staff principals to quickly and logically assess information and provide sound recommendations to the commander."[10] Yet in describing the procedures to be used, it sounds very much like FM 101-5. In fact, chapter 6 seems to acknowledge this as it offers an "abbreviated command estimate" and even this chapter offers no shortcuts to the standard five step estimate process.[11]

In 1988 the CGSC also produced a draft training circular, TC 101-5, Staff Techniques and Procedures to Support Command and Control, as an "abbreviated reference book for commanders and staff officers."[12] The October 1988 draft of TC 101-5 was much larger than the 1984 FM 101-5 and contained separate "how to" appendices on course of action development, war-gaming, and decision matrices. This document suggests a current tendency to further elaborate and prescribe the estimate process which had remained essentially unchanged for the past two decades.

There is, however, at the same time considerable concern about the

applicability of the estimate process on the modern battlefield. Indicative of this concern is the recent establishment of a CGSC working group to revisit the estimate process and determine what should go into the new edition of FM 101-5.

There is currently a greater knowledge of the tenants of the estimate process among Army command staffs than ever before. The creation of the Combined Arms Service and Staff School (CAS3) within the CGSC has resulted in the exposure of virtually every captain in the Army to the estimate process. At least equally important is the creation of the School for Advanced Military Studies (SAMS) within the CGSC. SAMS students undergo an intense year of training in large unit tactics with emphasis on application of the estimate process. Following graduation, many of them are assigned as plans officers in divisions and corps. The author has observed a noticeable increase in conscientious attempts to apply the estimate process in division field exercises where SAMS graduates are involved.

The Problem

The difficulties in applying a complete estimate process under battlefield conditions has two dimensions: time pressure and natural inclination. These two combine under the stress of warfare to create a potent inhibitor to the systematic deductive procedures of the estimate process.

The rapid pace of modern warfare creates great time pressure on tactical decision making. Added to this are the sophistication of battlefield systems, the complexities of their interactions, the amount of information available and the size of staffs plus the certainty and ability of the enemy to deceive. It is hard to conceive of a more difficult decision making environment then that presented on the modern air-land battlefield.

Lynch sees the time pressure of warfare as requiring a trade-off between two types of errors in the decisionmaking process.[13] If the commander goes through the entire estimate process to arrive at an optimum decision, errors due to uncertainty can be minimized. But under time pressure this means less time to detail the plan and for subordinate units to prepare and rehearse, less time to synchronize the battlefield operating systems. If the commander shortcuts the estimate process then time might be available for synchronization but there is more uncertainty about the situation and whether or not an optimum course of action has been selected. It is the balancing of these two types of errors that characterizes tactical decisionmaking under time pressure.

There is also some evidence that the estimate process with its generation and comparison of multiple options is not representative of naturalistic decisionmaking. Given a choice, tactical decisionmakers seem to prefer to produce a single option and develop it until they are either satisfied that it will work or that it must be modified or disgarded and another one generated.[14] Klein calls this approach to decisionmaking "progressive deepening."[15]

It appears that time pressure reinforces this "natural tendency" to seek shortcuts in the decision making process. Leddo, et.al. found that

time stressed tactical decision makers spent a relatively smaller proportion of time resolving uncertainty than did non-time stressed decision makers.[16]

Few would argue that the estimate process is not an objective, thorough and usually effective way of reaching a sound military decision. The problem is transistioning it from the relatively relaxed atmosphere of the classroom to the stressful and typically time compressed environment of field operations.

Recommendation

It is doubtful that any single effort will optimize the decision-making process in the field. What is recommended here is rather a three-sided approach including training, techniques and aiding, each of which has many facets.

In the area of training we need to know more about what constitutes expertise in tactical decisionmaking. It seems reasonable that one way to shorten the estimate process is for the decision maker to already possess much of the inferential understanding that is gained from a thorough estimate. With the proper expertise, good decisions are possible by sampling to test the validity of the expert's hypotheses rather than having to perform an exhaustive analysis. Toward this end, the ARI Field Unit at Fort Leavenworth has recently undertaken a project to examine the knowledges and abilities that are most characteristic of rapid tactical decision making expertise. If these can be adequately identified and isolated it may be possible to train expertise more efficiently.

Current instruction in the estimate process at the CGSC and elsewhere is felt to be invaluable in exposing the student to the science of warfare and to objectivity in tactical decisionmaking. A concerted and consistent effort should be made throughout the period of instruction to gradually decrease the amount of time available to reach a decision and prepare an order. Instruction should include how to tailor the estimate process under time constraints. It may eventually be possible to develop computer assisted instruction to enhance battlefield inferential skills and practice rapid tactical decisionmaking with performance feedback.

In the area of techniques current doctrine suggests that the primary means of reducing the estimate effort is by reducing the number of people involved. The command estimate described in ST 100-9 suggests that only the commander and his principal staff be involved.[17] FM 101-5 says that when time is critical "the commander may have to proceed through the decision-making process and issue orders based on his own knowledge of the situation without taking the time required to formally include the staff in the process."[18] Besides the two dimensions of the number of people involved and oral versus written orders, a third dimension to be considered must involve a reduction in the formal five step estimate process itself. An unpublished draft of FM 101-5 written in 1977 suggested that analysis and comparison of courses of action could occur simultaneously. Lynch suggests that the number of options considered be reduced under increasing time

constraints.[19] Other areas to consider reducing under time pressure are the amount of data seeking, the depth of wargaming and the objectivity of the process. How best to gracefully degrade the estimate process under increasing time pressure is fertile ground for research.

It is the author's personal opinion that the greatest gain in being able to apply the estimate process under battlefield conditions comes when the command thinks and plans as far ahead as is feasible. ARI has developed and implemented a method of assessing command and control performance that looks at the longevity and stability of planning as its primary measure. This method, the Army Command and Control Evaluation System (ACCES), views the command and control cycle as depicted in Figure 3. The command and control system monitors the current situation and develops understandings based upon the monitored data. If the system is functioning well the commander and his staff are forecasting or projecting the situation forward in time based upon these understandings to discover possible future problems and opportunities. When required the commander and staff begin the estimate process, formulating possible solutions and wargaming them to predict the events that must occur for them to succeed and the branches that might be required if certain anomalies occur. It is the two activities of forecasting and predicting that make a command proactive and give it sufficient lead time to make adequate use of the estimate process. It is failure to forecast that is the major cause of time pressure under battlefield conditions. It is the failure to predict or wargame through courses of action that makes forecasting such a difficult task because there is no "script" that tells the executor what to look for. It is the failure to do both of these that places the command in a reactive mode.
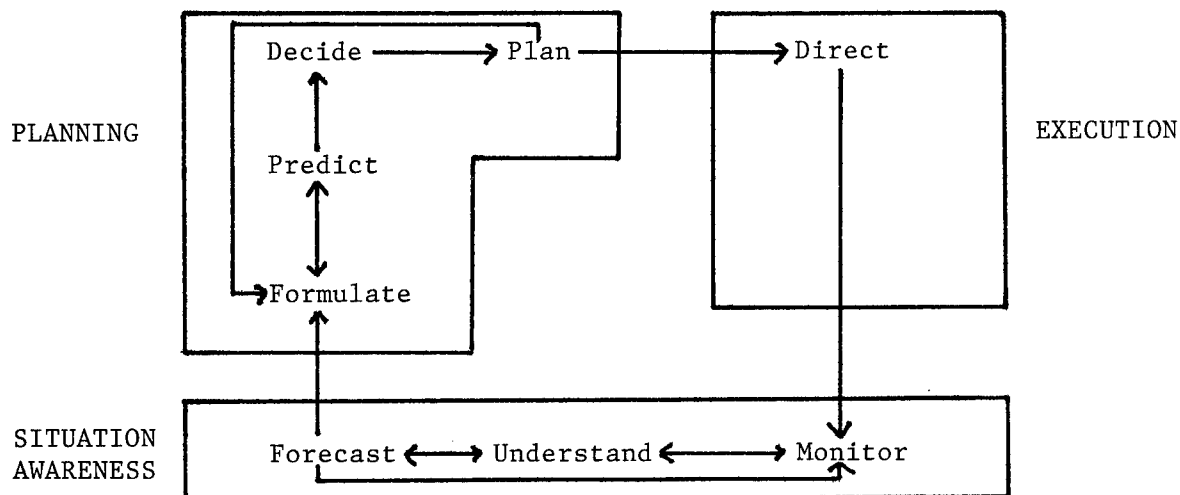

Figure 3. The command and control cycle.

The third area for improving estimate performance in the field is computer aiding. There are many planning aids being developed such as ALBM and Brigade Planner that are intended to help in formulating courses of action and wargaming and comparing them. ARI, Fort Leavenworth is experimenting with an aid that guides the decisionmaker through the estimate process.[20]

Computer aids can provide valuable supports to the tactical decisionmaker. Among these are computational support, recordkeeping, memory aiding, improved battlefield visualization, and improving the recognition of relationships among information items. For aids to be useable in field applications, however, they must not only be "user friendly", but they must truly make the estimate process doable under realistic time constraints. If they do not perform within the time available they cannot expect to fare any better than the manual estimate process.

In summary, transistioning the estimate process from the classroom to full consideration in the field requires training, adaptation to field conditions, and aiding. No one approach will be adequate by itself.

Endnotes

1. FM 101-5 (1984) Staff organization and operations, Washington, DC: Headquarters, Department of the Army.

2. Noted in, Lynch, Timothy D., Major, USA (1990) Problem-solving under time constraints: alternatives for the commander's estimate. Master of Military Arts and Sciences thesis, US Army Command and General Staff College, Fort Leavenworth, Kansas. Appendix B of that document contains an extract of Napoleon's written estimate.

3. Ibid, pages 10 and 11.

4. Ibid, page 11.

5. Field service regulations, United States Army, 1910, Washington DC: War Department, 1910. The paragraph appears under Article III, paragraph 75 on page 59.

6. Michel, Rex (in publication) Historical development of the estimate of the situation. ARI Technical Report.

7. Quoted from FM 101-5 (1960). Staff organization and operations. Washington DC: Headquarters, Department of the Army, page 56.

8. Ibid (1982), page 5-11.

9. Ibid (1984), page E-6.

10. Student Text 100-9 (1989). The command estimate. U.S. Army Command and General Staff College, Fort Leavenworth, Kansas, page 1-2.

11. Ibid, pages 6-1 and 6-2.

12. TC 101-5 (draft)(1988). Staff techniques and procedures to support command and control. US Army Command and General Staff College, October, 1988. Page i.

13. Lynch (1990), pages 35 and 36.

14. Examples of these types of findings are contained in: Michel, Rex R. and Riedel, Sharon L. (1988). Effects of expertise and cognitive style on information use in tactical decision making. ARI Technical Report 806, US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Thordsen, Galushka, Klein, Young and Brezovic (1990). A knowledge elicitation study of military planning. ARI Technical Report 876, US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

15. Klein, Gary (1989). Strategies of decision making. Military Review, 19, 5, May 1989, pages 56-64.

16. Leddo, Chennis, Cohen and Marvin (in publication) The influence of uncertainty and time stress on decision making. ARI Research Note.

17. ST 100-9, paragraph 1-4.

18. FM 101-5 (1984), page 5-5.

19. Lynch (1990), pages 31 and 32.

20. Fallesen, Jon, Michel, Rex and Carter, Charles (1989) Analysis of tactical courses of action using structured procedures and automated aids. In Proceedings of the Twenty-Eighth Army Operations Research Symposium, October, 1989.

MIDDLEBROOK

Immunological Evaluation of Phospholipase A_2 Neurotoxins (U)

Dr. John L. Middlebrook
U.S. Army Medical Research Institute of Infectious Diseases
Frederick MD 21701-5011

Toxins have long been recognized as biowarfare threats to U.S. forces. Before the advent of molecular genetics, the threat toxins were few in number, due to the practical considerations of being able to produce sufficient quantities for weaponization. Thus, the main toxins of concern were botulinum toxin(s), staphylococcal enterotoxin(s) and ricin. Now, with the highly practical research technology of molecular genetics, it is theoretically possible to clone and express any protein toxin in large amounts. This technical advance makes the list of potential threat toxins very long. Although botulinum toxin remains the most toxic substance known to science, there are many types of toxins in nature that are more toxic than either ricin or the staphylococcal enterotoxins. Thus, nations or groups unfriendly to the U.S. or its allies could select one or more of these novel protein toxins and, by the application of molecular genetics, produce a potent weapon. This new and long list of potential toxin threats presents a significant new challenge to the U.S. defense program involving development of medical countermeasures to biowarfare agents. We simply do not have the resources to develop antidotes or vaccines against every possible toxin. Clearly, our research efforts must be structured so as to produce products and solutions that will work against the broadest possible spectrum of toxins.

Phospholipase A_2 neurotoxins are produced from snake species found worldwide. At least 14 distinct phospholipase A_2 neurotoxins have been reported thus far, and it is likely that more will be discovered. Of all the protein toxins known, only microbial toxins (such as botulinum and diphtheria) are more toxic to animals and humans. As a class, phospholipase A_2 neurotoxins are the most toxic proteins found in snake venoms. Much evidence suggests that these toxins are the major cause of death due to the bites of snakes whose venoms contain this class of neurotoxin. These toxins are lethal by the aerosol route, and several have

MIDDLEBROOK

cloned. In the view of many, phospholipase A_2 neurotoxins are leading candidates as toxin biowarfare weapons, and the development of protective measures is warranted.

TABLE 1: PROPERTIES OF THE PHOSPHOLIPASE A_2 NEUROTOXINS

TOXIN	SNAKE	MOLE WT	# CHAIN	LD50(μg/Kg)
taipoxin	taipan	40k	3	2
beta bungarotoxin	banded krait	20k	2	20
crotoxin	S.A. rattlesnake	20k	2	50
Mojave toxin	Mojave rattlesnake	20k	2	30
concolor toxin	midget-faded rattlesnake	20k	2	45
notexin	tiger snake	14k	1	17
ammodytoxin	European viper	14k	1	20
C. atrox phospholipase	western diamondback rattlesnake	14k	1	0
C. vegrandis phospholipase	S. American rattlesnake	14k	1	50
N. a. atra phospholipase	Formosan cobra	14k	1	5000
caudoxin	horned puff adder	14k	1	180
textilotoxin	brown snake	80k	4	1
pseudexin	black snake	14k	1	1250

MIDDLEBROOK

The phospholipase A_2 neurotoxins have a variety of structures (Table 1) ranging from the complex heteropentamer, textilotoxin (MW 60,000) to the simple monomer, notexin (MW 14,000). Likewise, the lethal potencies of these toxins ranges over three orders of magnitude, from 1 to 1250 µg/kg in mice. The literature covering the immunology of these toxins is not extensive, but several studies have been carried out[1]. By and large, the conclusions reached by the investigators state that few or no immunological similarities or relationships exist among this class of toxins. Indeed, a recent review[1] by an authority in the field concluded "However, and this is the most striking observation arising from these immunological studies, the antibodies obtained are exceedingly specific for each toxin against which they are produced." While consistent with the data from several studies, this conclusion was nonetheless surprising since the primary sequences for several of the toxins were known[2] and these were clearly homologous in certain regions. Because of this apparent contradiction, we decided to reinvestigate the question of immunological relatedness of the phospholipase A_2 neurotoxins.

We began by purifying 13 different phospholipase A_2 neurotoxins and enzymes from their venom sources (Table 1). Although purification schemes were published, we improved upon each of these schemes by applying the technology of FPLC chromatography (Pharmacia Corporation). This approach, which is intermediate between open column and high pressure liquid chromatography, gives superior resolution of moderate to large quantities of venoms. As judged by SDS gel electrophoresis, each of our products was highly pure (~95%), and a mouse lethality assay was performed to assure that the toxins were fully active. In several instances, our products were run through 5-8 cycles on an amino acid sequencer to assure authenticity (assistance of Dr. James Schmidt, this Institute).

Following purification, each toxin was used to immunize rabbits. There was nothing unusual about the immunization schedule or technique. The rabbits received doses of 10-100 µg of toxin (depending upon the toxicity of the individual toxin) and Freund's adjuvant, intramuscularly at two week intervals. The toxins were not inactivated or toxoided in any way. In 6-8 weeks, the rabbits appeared to have a high circulating level of antibody. Each rabbit was bled and boosted in succession, and a good supply of antisera was stockpiled.

We asked several basic questions using the rabbit antisera. First, we wanted to know whether or not these toxins are good antigens. The answer was determined by the technique of ELISA. After binding each toxin to the wells of microtiter plates,

MIDDLEBROOK

we added serial dilutions of rabbit antisera, followed by an indicator second antibody. Figure 1 shows the representative examples of data we obtained.[3] Easily measured color reactions were observed at dilutions out to 1:100,000. Each of the rabbit antisera produced exhibited these strong reactions, and we were surprised to find titers that were approximately the same for each toxin and its homologous antiserum. We anticipated significant variability in the responses of different animals to different toxins, but such was not the case.
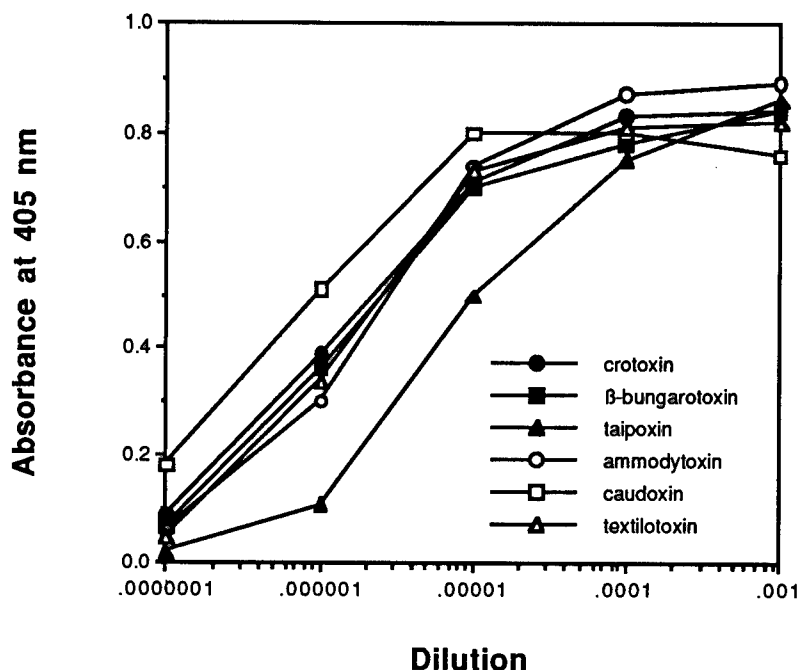
Fig. 1. Homologous ELISA titers of the phospholipase A$_2$ neurotoxins. Plates were coated with toxins as indicated, then incubated with the specified dilutions of homologous rabbit antisera, and finally with goat anti-rabbit antisera conjugated to peroxidase. Substrate was added and absorbances read after 30 min.

Next, we sought to determine the degree of cross-reactivity that exists among the toxins and their heterologous antisera. This was accomplished with the same ELISA technique mentioned above. A complete matrix of possible cross- reactions was investigated. Since these data have been published[3], they will not be reproduced here. Basically, we found many instances where antiserum against a

MIDDLEBROOK

given phospholipase A_2 neurotoxin was capable of reacting strongly with other phospholipase A_2 neurotoxins. We determined that these toxins fall into two main seroclasses, the elapid toxins in one and the crotalid and viperid toxins in the other. Antiserum against one member of a class will generally react quite strongly with all the other members. Thus, by selecting only two antisera (one from each class), we should have a "reagent" capable of detecting all phospholipase A_2 neurotoxins. These surprising results were somewhat at odds with the literature cited above, but they were checked very carefully and are reproducible.

As luck would have it, an opportunity arose to test our hypothesis. Agkistrotoxin[4] is a recently discovered phospholipase A_2 neurotoxin from the Chinese pit viper, *Agkistrodon halys pallas*. At the time we produced our
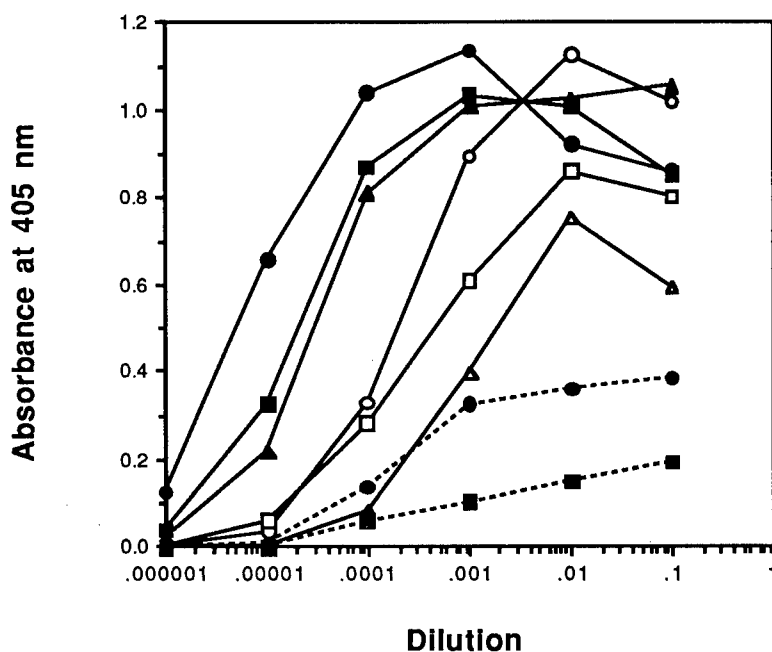
Fig. 2. Agkistrotoxin ELISA. A plate was coated with agkistrotoxin, then incubated with the specified dilutions of homologous rabbit antisera and finally goat anti-rabbit antisera conjugated to peroxidase. Substrate was added and absorbances read after 30 min. (●), crotoxin antiserum; (■), Mojave toxin antiserum; (▲), concolor toxin antiserum; (O), caudoxin antiserum; (□), ammodytoxin; (△), ß-bungarotoxin antiserum; (--●--), notexin antiserum; (--■--), normal rabbit serum.

MIDDLEBROOK

rabbit antisera and performed the experiments delineated above, this toxin was not available in the western world. We have now obtained limited quantities of this toxin and tested it with our standard ELISA (Figure 2). Clearly, several of the antisera against other phospholipase A_2 neurotoxins react strongly with agkistrotoxin and are quite suitable as detection antibodies. Thus, the hypothesis that our rabbit antisera should react with any phospholipase A_2 neurotoxin was been born out in our first test case.

Our next step was to analyze these same antisera for their capabilities to neutralize and cross-neutralize the lethality due to the phospholipase A_2 neurotoxins. These tests were performed by mixing antiserum with toxin, then injecting the mixture into a mouse. By using increasing amounts of toxin with a

FIG. 3. Neutralization of phospholipase A_2 neurotoxins by antiserum to textilotoxin. Textilotoxin antiserum (0.1 ml) was mixed with the specified dose of toxin, incubated at room temperature for 30 min. and injected ip in mice. Animals were observed for 3 days, and deaths recorded.
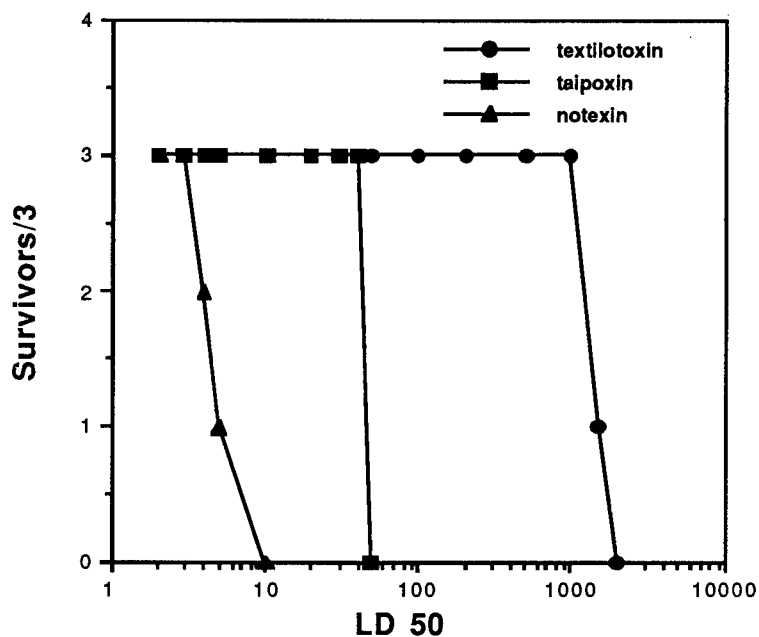
MIDDLEBROOK

constant dose of antiserum, we could titer the protective efficacy of each antiserum against a given toxin. In all cases, we found that homologous antiserum would protect mice from the lethal effects of that phospholipase A_2 neurotoxin. An example of such determinations is shown in Fig 3. Antiserum to textilotoxin protected completely from 1000 LD_{50}s of textilotoxin, but not measurably from 2000 LD_{50}s. The same antiserum completely protected from 50 LD_{50}s of taipoxin and 3 LD_{50}s of notexin. Several other elapid phospholipase A_2 toxins were also tested with negative results (data not shown). Thus, it appears that textilotoxin, taipoxin and notexin have a similar epitope or epitopes which are critical for lethality. When one calculates the number of molecules of toxin that are neutralized (a function of molecular weight and potency), the differences are not as striking. It was determined that an equivalent amount of textilotoxin antiserum would neutralize 10 molecules of textilotoxin, 3 molecules of notexin and 2 molecules of taipoxin. This calculation suggests that the neutralizing epitopes shared by these molecules are rather similar in structure. Similar immunological cross-protections for all the phospholipase A_2 neurotoxins listed above have been defined by the same experimental protocol as that used to obtain the data in Fig. 3. Using this information as a base, we are now attempting to identify the structure of these common neutralizing epitopes in hopes of using that information to prepare and test possible vaccines.

Although antiserum is a useful and valuable research tool, producing sufficient quantities for fielding in identification kits is difficult and expensive. The antibody reagent of choice for this task is a monoclonal antibody. Thus, we embarked on a series of studies designed to generate monoclonal antibodies against the phospholipase A_2 neurotoxins. We were successful in producing over 50 cell lines against 7 different neurotoxins. The first monoclonal antibodies we obtained were against the crotalid neurotoxin, crotoxin[5]. A total of 12 lines were isolated; however, only four of these produced antibody when grown as ascites tumors. Fortunately, any one of these four would serve well for a detection/identification kit. Representative reactions are shown by the data in Table 2. As can be seen, the cross-reactivities of two of the monoclonal antibodies are quite strong. There are essentially no differences in the reaction of #11 with crotoxin, concolor toxin or Mojave toxin. Significant, but small differences are seen with antibody #5 and the same toxins. Both antibodies react less efficiently with vegrandis toxin. However, since the absolute ELISA titer vs. crotoxin is very high (1-10 ng of antibody produces intense color), this lower cross-reactivity with vegrandis toxin can be easily overcome by using slightly higher quantities of antibody. These crotoxin monoclonal antibodies had no detectable reactions with any of the viperid or

MIDDLEBROOK

Table 2: Cross-reactions of Crotoxin Monoclonal Antibodies

Toxin	Antibody required*	
	line #11	line #5
crotoxin	1.0	1.0
Mojave toxin	1.3	2.7
concolor toxin	1.6	2.6
vegrandis toxin	4.0	8.1

* Values represent the amount of antibody required to produce a standard absorbance value in an ELISA.

elapid toxins, so it was necessary to produce another set of monoclonal antibodies that would recognize these other two classes of phospholipase A_2 neurotoxins.

Pseudexin[6] is a phospholipase A_2 neurotoxin found in the venom of the Australian red-bellied black snake, *Pseudechis porphyriacus*. Recent work has shown that pseudexin is really a mixture of three isoenzymes, two of which are toxic[7]. Since work with rabbit polyclonal antibody demonstrated a large number of cross-reactions of snake phospholipase A_2 neurotoxins with antisera to pseudexin[3], we chose to examine the properties of 15 pseudexin-derived monoclonal antibodies. We began with the ELISA reactions of the monoclonal antibodies to pseudexin and the other phospholipase A_2 neurotoxins. In a majority of instances, a given monoclonal antibody reacted only with pseudexin(s) and one or two additional phospholipase A_2 neurotoxins. However, three of the monoclonal antibodies reacted with all of the phospholipase A_2 neurotoxins we studied. An example of this type of behavior is shown in Figure 4. Monoclonal antibody #14 reacted strongly with all the elapid neurotoxins tested and moderately with the crotalid and viperid toxins. When compared side-by-side (data not shown), the reactions of #14 with elapid toxins were as intense as the reactions of crotoxin-derived #11 with the crotalid toxins. Thus, the combination of these two monoclonal antibodies is a highly sensitive reagent suitable for the detection of low amounts of elapid or crotalid toxins. While the use of higher levels of #14 would enable one to also detect viperid toxins, we have recently obtained a caudoxin-derived monoclonal antibody that, when added to monoclonal antibodies #14 and #11, will give us a sensitive detection reagent recognizing all phospholipase A_2 neurotoxins.
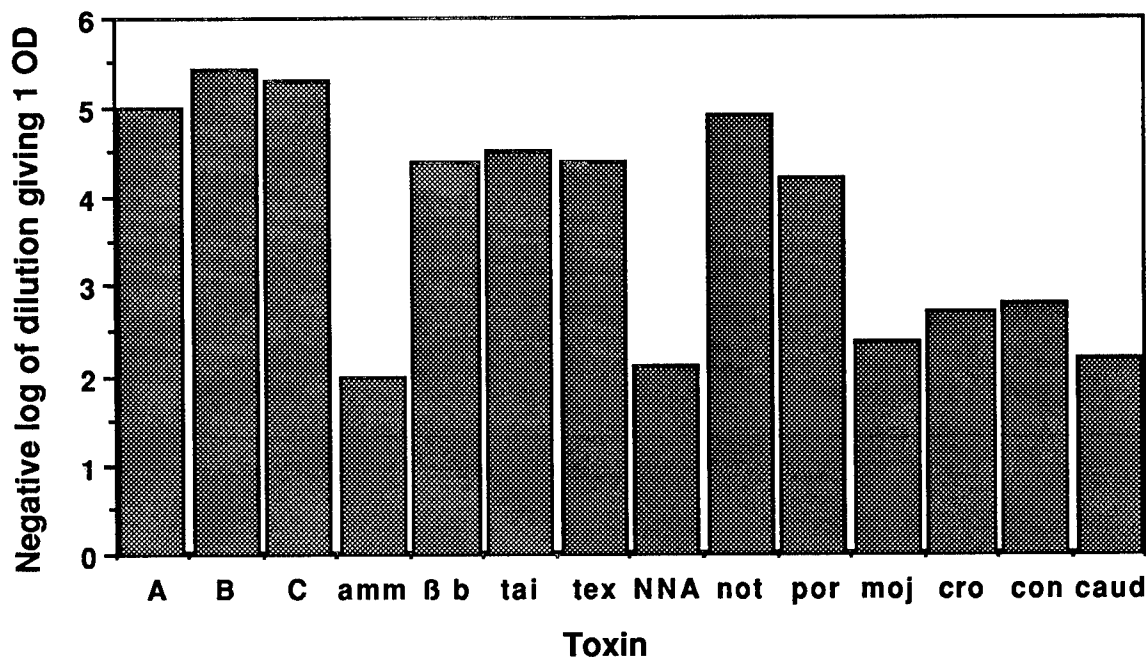
Fig 4. Immunological cross-reactions of pseudexin-derived mono-
clonal antibody with phospholipase A_2 neurotoxins. Plates were coated
with various toxins, then serial dilutions of monoclonal antibody #14 were added.
After washing, rabbit anti-mouse antibody conjugated to peroxidase was added
followed by substrate. The optical density (OD) at 405 nm was read after 30 min.,
and the log of that dilution producing 1 OD was determined graphically. Toxins
were as follows: A - pseudexin A; B - pseudexin B; C - pseudexin C; amm -
ammodytoxin; ß b - ß-bungarotoxin; tai - taipoxin; tex - textilotoxin; NNA - *Naja
naja atra* acidic phospholipase A_2; not - notexin, por - porcine pancreatic
phospholipase A_2; moj - Mojave toxin: cro - crotoxin; con - concolor toxin; caud -
caudoxin.

One of the most important goals of our research efforts is to develop protective
drugs or vaccines against the toxins we study. An exciting possibility was that
some of the monoclonal antibodies we obtained would neutralize the toxin(s) of
concern and could be used as protective therapy. As reported earlier[5], we found
that one (but only one) of the crotoxin-derived monoclonal antibodies would
neutralize crotoxin, Mojave toxin and, to a lesser degree, concolor toxin. The

MIDDLEBROOK

pseudexin-derived monoclonal antibodies were also tested with the resultant finding that three of the fifteen neutralized pseudexin, but no other phospholipase A_2 neurotoxin (J. L. Middlebrook, unpublished data). Thus, while some neutralizations are observed, the spectrum of such is not wide enough to suggest that the monoclonal antibodies we have would serve as useful therapeutic agents for Army purposes.

While the monoclonal antibodies were not suitable for clinical purposes, it is possible to use them as probes to define the structure of the neutralizing epitopes on the toxins. This we have done by the use of research kits provided by Cambridge Research (Cambridge, MA.). In the first approach, we asked whether the monoclonal antibodies recognize linear epitopes. This was accomplished by synthesizing consecutive, partially overlapping peptides corresponding to the entire sequence of the toxins that were neutralized. Then, using the peptides as capture antigens in an ELISA, we tested each monoclonal antibody to see if it would recognize and bind to any of the synthesized peptides. With both the crotoxin-derived monoclonal antibody (Dr. I. Kaiser, personal communication) and the pseudexin-derived monoclonal antibodies (J. L. Middlebrook and B. Styles, unpublished observations), no positive ELISA reactions were observed. These results indicated that our monoclonal antibodies do not recognize linear epitopes and, therefore, must bind to conformational epitopes defined by the three dimensional structure of the toxins. In hopes of gaining insights into the nature of the three dimensional epitopes, we are now using the "mimiotope" approach specified by the Cambridge Research kits. This work is currently in progress, but no structures have yet been identified.

In summary, the work described above has produced both a product and a lesson. The product is represented by the monoclonal antibodies, which can be used immediately as an ELISA-based detection system for phospholipase A_2 neurotoxins. The lesson may, in the long term, be more important. We have learned not to be intimidated or overly concerned with a long list of potential biowarfare toxins. We no longer assume that it will be necessary to develop a like number of detection reagents and therapies, even in the face of literature reports which might suggest so. That is not to minimize the gravity of the toxin biowarfare threat which faces us, but rather to assert generic approaches exist and should be explored.

MIDDLEBROOK

REFERENCES

1. A. Menez, Pharmac. Ther. 30, 91 (1985).
2. D. Mebs, List of Biologically Active Compounds from Snake Venoms, Frankfurt University Press (1985).
3. J.L. Middlebrook and I. I. Kaiser, Toxicon 27, 965 (1989).
4. M. Jiang, J. Haggblad, E. Heilbronn, B. Rydqvist, and D. Eaker, Toxicon 25, 785 (1987).
5. I.I. Kaiser and J.L. Middlebrook, Toxicon 26, 855 (1988).
6. G.T. Vaughan, T.B. Sculley and L. R. Tirrell, Toxicon 19, 95 (1981).
7. J.J. Schmidt and J.L. Middlebrook, Toxicon 27, 805 (1989).

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Circumventing Antimalarial Drug Resistance: Theoretical and Clinical Implications.

WILBUR K. MILHOUS*, LTC, MSC, DENNIS E. KYLE, CPT, MSC
RODGER K. MARTIN, CPT, MSC, MAX GROGL, CPT(P), MSC,
AYOADE M. J. ODUOLA, Dr., RICHARD N. ROSSAN, Dr.
AND BRIAN G. SCHUSTER, COL, MC
Walter Reed Army Institute of Research, Washington, D.C.
Gorgas Memorial Laboratory, Panama

Military campaigns have been plagued throughout recorded history by the most debilitating and fatal form of human malaria caused by the protozoan parasite _Plasmodium falciparum_. Drug resistant malaria was first encountered during the Vietnam conflict and was the major cause of medical disability. Resistance to available drugs is increasing both in severity and prevalence (1) and is present in many areas of strategic importance. Clinical resistance to the Army's new antimalarial drug mefloquine has been observed in Thailand (2) and naturally occurring foci of mefloquine resistance have been reported in West Africa (3). The development of new antimalarial drugs has been a slow, expensive and complex process that has not only been confounded by the parasite's ability to develop resistance to existing and developmental drugs, but also by the biological diversity of the parasite's complex life cycle. Recognizing the limited clinical utility of existing antimalarial drugs and potential for development of multiple drug resistance, it has been critical to establish a drug development program with long term strategies to predict, quantitate and circumvent the drug resistance phenomenon.

In order to identify novel drug substances, numerous approaches and strategies have been employed to design and synthesize new antimalarial drug candidates. In the history of the U.S. Army Antimalarial Drug Development Program, these approaches have traditionally included empirical screening, screening of plant extracts from traditional medicine, lead-directed synthesis of compounds known to have antimalarial, antimicrobial or antitumor activities, and selective targeting of specific parasite enzymes. An essential component of the antimalarial drug discovery process has been the use of _in vitro_ techniques to evaluate the potential antimalarial activities of new drugs. Radioisotopic methods for _in vitro_ drug screening were intially developed in 1979 at the Walter Reed Army Institute of Research (4) and have been subsequently automated and optimized to examine the intrinsic activities of antimetabolites and antibiotics (5) and to perform quantitative drug interaction studies (6).

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

In vitro screening has served as the foudation of the drug discovery process. The Walter Reed Army Institute of Research has utilized well characterized clones of falciparum malaria parasites which have been derived by direct visualization and single cell micro-manipulation (7). A parasite clone designated as Indochina W-2 (derived from CDC Indochina III) is resistant to the the antimalarial drugs chloroquine, pyrimethamine, sulfadoxine and quinine but susceptible to mefloquine. The Sierra Leone D-6 clone (derived from CDC Sierra Leone) is a naturally occuring mefloquine-resistant clone which is susceptible to other antimalarial drugs. Historical data on the stability of the concentration-response profiles of these clones used in antimalarial drug screening at the Walter Reed Army Institute of Research and World Health Organization are presented in Table 1.

Table 1. Mean in vitro fifty percent inhibitory concentrations (IC_{50} in ng/ml) and calculated resistance indices of Plasmodium falciparum Indochina W-2 and Sierra Leone D-6 Clones against standard antimalarial drugs during the period 1985 through 1990.

Antimalarial Drugs	W-2 Clone	D-6 Clone	Resistance Indices
Pyrimethamine	43.11	00.07	616
Sulfadoxine	15,934	22.59	705
Chloroquine	51.20	02.19	23.4
Quinine	50.69	09.99	5.07
Mefloquine	01.22	07.27	5.95

This stable and well defined pattern of drug susceptibility makes these clones extremely useful for evaluating candidate drugs of diverse chemical classes. These clones are also used as simultaneous controls for characterizing patient parasite isolates as part of an intensified surveillance program at the Walter Reed Army Institute of Research to which critically assesses the case reports of drug-resistant falciparum malaria from different geographic regions. The use and comparison of isolates from patients with known clinical outcomes has been instrumental in our ability to predict potential patterns of cross-resistance among new antimalarial drugs in various stages of preclinical development. These data not only provide valuable current epidemiological data to insure that U.S. soldiers are adequately protected in endemic regions, but also insure that the process for design and selection of new drugs in being guided by the most current biological data. Selected isolates with unique drug susceptibility patterns have undergone molecular characterization in our own attempt to decode these patterns and investigate mechanisms of resistance.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

A striking example of the use of these techniques has been in the evaluation of parasites with varying levels of susceptiblity to the antifolate antimalarials pyrimethamine and cycloguanil (see Table 2). Clinical reports have suggested that parasites that were resistant to one drug might be successfully treated with the other drug and vice versa. As a result, the specfic enzyme target, dihydrofolate reductase (DHFR), has been analyzed from different parasites to reveal the structural basis for altered binding affinity and differential drug susceptibility (8). Parasites which had amino acid point mutations which resulted in substitutions of alanine-16 to valine-16 and from serine-108 to threonine-108 were found to be resistant to cycloguanil. A single point mutation (aspargine-108) was found to confer resistance to pyrimethamine, but only resulted in modest decreases in susceptibility to cycloguanil. Significant cross resistance to both drugs was found to occur in parasites having point mutations which included a change from serine-108 to asparagine-108 and from isoleucine-164 to leucine-164 (see Fig. 1). A structural analog of cycloguanil, WR99210 (2',4',5'-trichlorophenoxy substituted triazine) was found to be remarkably active against all of the clones tested with an IC_{50} of less than 0.01 ng/ml. This broad spectrum of .enhanced activity of WR99210 (Table 2) and lack of cross resistance to other antifolate antimalarials provide valuable insight for lead directed synthesis of future drugs targeted at the parasite's DHFR.
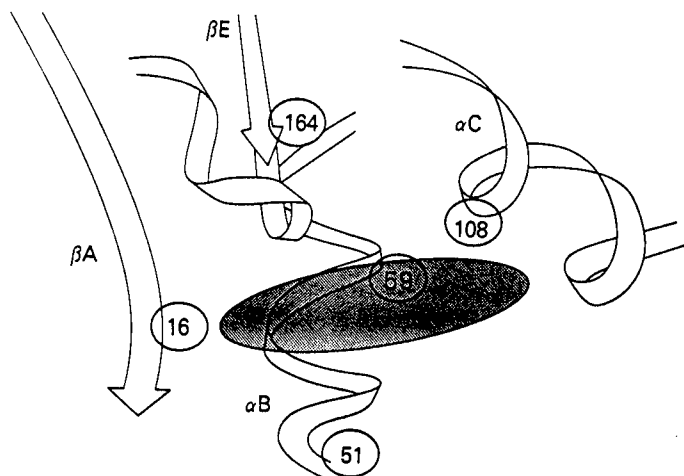
Fig 1. Schematic representation of the point mutations in the active site of falciparum malaria DHFR enzyme. Point mutations which result in cycloguanil resistance are associated with valine substitutions at position 16 and threonine at positiion 108. Pyrimethamine resistance is associated with aspargine substitution at position 108. Parasites with additionals mutations at positions 51, 59 and 164 have enhanced resistance to both pyrimethamine and cycloguanil but not WR 99210. Peterson, D.S., Milhous, W.K. and Wellems, T.E. (1990). Copyright PNAS.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Table 2.   In vitro fifty percent inhibitory concentrations (ng/ml) of pyrimethamine, cycloguanil and a cycloguanil analog, WR99210, against falciparum malaria parasites with differential susceptibility to both pyrimethamine and/or cycloguanil.   Note that WR99210 is remarkably active against even the pryimethamine and/or cycloguanil resistant strains (R = resistant).

Parasite Strain	Pyrimethamine		Cycloguanil		WR99210
D-6 Sierra Leone	00.03		00.03		0.002
FCR3 Gambia	00.56		53.8	R	0.002
FCB Columbia	00.41		30.9	R	0.003
HB3 Honduras	21.5	R	00.45		0.003
Itd12 Brazil	88.1	R	00.41		0.002
W2 Indochina	33.9	R	00.68		0.002
V1S Vietnam	83.3	R	24.40	R	0.002

DISCOVERY OF THE MULTIPLE DRUG RESISTANCE PHENOMENON IN MALARIA PARASITES

At the Walter Reed Army Institute of Research, experimental evidence for a multiple drug resistance (MDR) phenomenon in falciparum malaria began to emerge in 1985 when clincial trials were performed in human volunteers to determine the minimum effective dose of a new candidate pyridinecarbinol, enpiroline (9).   In one subject, after a single 500 mg dose of this drug, parasitemia recurred after three weeks.   The patient was then treated with the quinoline methanol, mefloquine, and subsequently failed.   In vitro susceptibility testing of the primary inoculum (Vietnam Smith isolate) and comparison with the second and third recrudescence cultures demonstrated a two-fold decrease in susceptibility of the parasites to enpiroline with a collateral decrease in susceptibility to mefloquine and the phenanthrene-carbinol halofantrine.   Similar experiments were conducted to induce resistance in cloned parasites to determine if the acquisition of resistance was a function of population selection or mutation.   When parasite clones were subjected to step-wise increasing concentrations of enpiroline over a three month period, a five-fold increase in resistance was obtained.   Drug pressure was removed for a period of three months and the resistance remained stable.   This five-fold increase in resistance to enpiroline was accompanied by a five-fold concurrent increase in resistance to mefloquine and halofantrine (6), drug of different chemical classes.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Results from subsequent experiments on the induction of resistance to halofantrine and the unique sesquiterpene lactone, artemisinin, suggested a similar phenomenon. The acquisition of drug resistance appeared to be independent of the chemical class of drug to which the parasites were exposed. More importantly, the quantitative differences between resistant vs. susceptible parasites was apparently much more subtle (2-10 fold) than had been observed (500-fold) with the antifolate antimalarials.

A similar multiple drug resistance scenario had been noted in cancer chemotherapy. Tumors in patients treated with one drug and developing resistance apparently also became resistant to drugs of different chemical classes. In vitro drug transport studies revealed that the level of drug resistance was related to the tumor cell's ability to efflux the drug out of the cells (10). Certain calcium antagonists, such as verapamil (11) were found to reverse this process in vitro. In vitro studies with malaria parasites were subsequently conducted whereby WRAIR scientists (12) demonstrated that the calcium channel blocking drug verapamil effectively reversed resistance to chloroquine in the chloroquine-resistant parasite clone but had no effect on chloroquine-susceptible parasites (see Fig. 2).
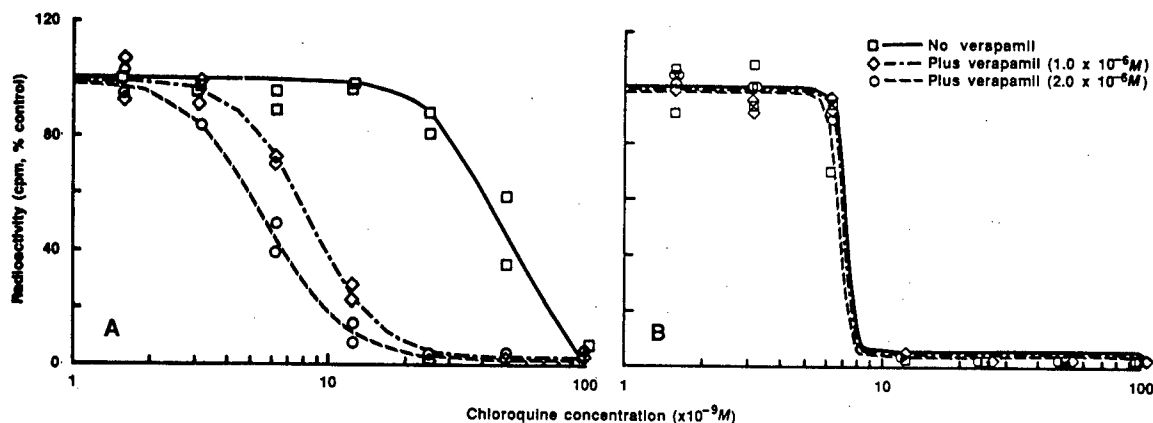

Figure 2. Dose response curves of chloroquine-resistant P. falciparum clone (W-2) from Indochina and (B) chloroquine-sensitive P. falciparum clone (D-6) from West Africa in the absence or presence of a constant concentration of verapamil. Martin, S.K., Oduola, A.M.J., and Milhous, W.K. (1987). Copyright AAAS.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Chloroquine resistant clones and patient isolates obtained from Indochina, East and West Africa, and South and Central America have undergone evaluation. These data suggest the mechanism of resistance to chloroquine and to its active metabolite, desethylchloroquine, is similiar in each of these isolates. The calcium antagonists including diltiazem, nifedipine, chlorpromazine and their various analogs have also been found to reverse resistance to chloroquine and enhance the intrinsic activities of quinine and quinidine (13). A piperidine analog, penfluridol, has been found to reverse resistance to the antimalarials mefloquine, halofantrine and artemisinin (14) . These studies have led to the new rationale for the development of "resistance modulator" drugs targeted specifically toward the parasite's mechanism of drug resistance.

Antimalarial drug resistance in P. falciparum thus appears to be analogous to multi-drug resistance (MDR) in neoplastic cells where the induction of resistance with one drug confers resistance to other structurally and functionally unrelated drugs. The mechanism of MDR in tumor cells is mediated by enhanced outward transport of drug which reduces the intracellular accumulation to sublethal levels. Amplification of a MDR gene has been observed (15,16) in malaria that is closely related to a mammalian MDR gene that encodes a P-glycoprotein which reduces intra-cellular accumulation of various drugs. Using the polymerase chain reaction (17), the P. falciparum gene for MDR has been sequenced from isolates from five different geographic regions and correctly predicted the parasite phenotype (susceptible vs. resistant) in forty-six of forty-eight clinical isolates.

With the development of an in vitro efflux model for falciparum malaria, studies (18) have also demonstrated that this drug transport process is a mechanism of resistance to chloroquine in malaria. If pharmacologically active drugs can be designed which will selectively inhibit drug efflux without significant toxicity to the host, this may prove to be a valuable new approach to the chemotherapy of multiple-drug resistant malaria.

EVALUATION OF CANDIDATE RESISTANCE MODULATOR COMPOUNDS

Candidate resistance modulator compounds are systematically evaluated in vitro utilizing two methods. Fifty percent inhibitory concentrations are determined for each drug alone and in combination at fixed ratios. The resulting data are plotted on conventional isobolograms to demonstrate additivity, synergism or antagonism. While the isobologram method is a good predictor of drug interactions, this method does not allow for quantitation of the resistance modulators. With the exception of the antihistamine drugs, ketotifin and cyproheptadine, most calcium antagonists have only weak intrinsic antimalarial activities (see Table 5).

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Evaluation of numerous candidate modulators has suggested that their mechanism of action is not directly related to their cytotoxic effect, but rather to their ability to block the physiological processes responsible for elimination of toxic drugs by the parasite. Therefore, drug combination studies are also conducted using constant sub-inhibitory concentrations of the candidate resistance modulator in the presence of varying concentrations of the antimalarial drug. Concentration response data obtained using this method (19) is used to construct a Response Modification Index (RMI) utilizing the following formula where A is the antimalarial drug and B is the candidate modulator drug:

$$RMI = IC_{50(A+B)} / IC_{50A}$$

A RMI of 1.0 would represent no change in intrinsic antimalarial effects, while an RMI of less than 1.0 would indicate potentiation. Thus the RMI provides quantitative evaluation of the interaction of two drugs where one drug (such as chloroquine) has intrinsic antimalarial activity and the other compound (such as verapamil) has little or no antimalarial activity (see results of testing in Tables 3 and 4).

Table 3. Effect of verapamil (1×10^{-6}M) or desipramine (1×10^{-7}M) on fifty percent inhibitory concentrations of chloroquine, quinine or quinidine against falciparum malaria clones. Note that verapamil or despiramine markedly reduces the IC_{50} (decreases the RMI) of chloroquine, quinine, and quinidine in the resistant strain, but has no effect on the chloroquine susceptible clone.

Antimalarial Drug	Drug Resistant Indochina Clone			Drug Susceptible Sierra Leone Clone		
	−	+VER	+DMI	−	+VER	+DMI
Chloroquine	161.9[1]	19.6 (.12)	47.7 (.29)[2]	4.8	4.7 (.96)	6.2 (1.3)
Quinine	94.2	32.0 (.34)	40.0 (.42)	29.5	33.9 (1.2)	36.4 (1.2)
Quinidine	33.1	7.4 (.22)	11.4 (.34)	10.8	11.2 (1.1)	10.8 (.97)

[1] IC_{50} (nM) without (−) and with (+) verapamil (VER) or desipramine (DMI)
[2] Response Modification Index (RMI)

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Table 4. Effect of penfluridol (2.5×10^{-7}M) on the response of P. falciparum to halofantrine, artemisinin and mefloquine. Note that penfluridol markely reduces the Response modification index (RMI) of the Nigeria ST and D-6 strains (mefloquine resistant strains), but not the W-2 strain which is mefloquine susceptible.

Parasite Strain	Halofantrine		Artemisinin		Mefloquine	
	−	+	−	+	−	+
Nigeria ST	4.8[1]	0.71	7.5	2.62	29.2	5.84
		(.15)[2]		(.35)		(.20)
D-6 Sierra Leone	4.3	0.95	7.26	2.9	26.9	6.99
		(.22)		(.40)		(.26)
W-2 Indochina	0.50	0.52	2.97	3.12	2.37	2.23
		(1.03)		(1.05)		(.94)

[1] IC_{50} (nM) with (+) and without (−) penfluridol (2.5×10^{-7}M)
[2] Response Modification Index (RMI)

Compounds screened using this method are subsequently selected for in vivo testing. A rodent model (20) using a chloroquine resistant strain has been developed, but the only model for human falciparum malaria is in the Aotus monkey. Monkeys are infected with the chloroquine resistant Vietnam Smith strain of Plasmodium falciparum (21). Each monkey is inoculated intravenously with 5×10^6 trophozoites which is an inoculum size which results in parasitemia on the fifth day post inoculation that can range fom 5,000 to 40,000/cm^3. At this time treatment is begun. Chloroquine alone or in combination with the candidate resistance modulator drug is administered orally for three or seven consecutive days. Untreated controls are also utilized. Chloroquine given at either three daily doses of 20 mg/kg or seven daily doses of 5-10 mg/kg is effective at curing infections with susceptible parasites; however, this treatment regimen is not effective against the chloroquine resistant Vietnam Smith strain. Blood films are prepared and exmained daily on all animals before and during treatment and for up to 100 days thereafter. Treated animals are observed twice daily for signs of drug toxicity, as evidenced by abnormal behavior, anorexia, diarrhea or vomiting. Necropies are performed on all animals that die.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

Several compounds (22,23) including verapamil, nifedipine, diltiazem, an analog of tiapamil, bepridil, desipramine (see Figure 3) and chlorpromazine have been evaluated in this simian model. Preliminary results with multiple dose regimens have suggested that some of the compounds have exhibited initial parasite suppression and clearance, but total cures have only been achieved after retreatments. Toxicity has also been noted with many of these drugs and the effect of these compounds on multi-drug transporter proteins found in normal cells (24) remains to be elucidated. Advanced pharmacokinetic studies are being performed to elucidate how one drug might alter the metabolism and pharmacodynamics of the second drug. When compared to most antimalarial drugs, most of the commonly used calcium antagonists also exhibit short plasma half-lives and limited oral bioavailability (see Table 5).
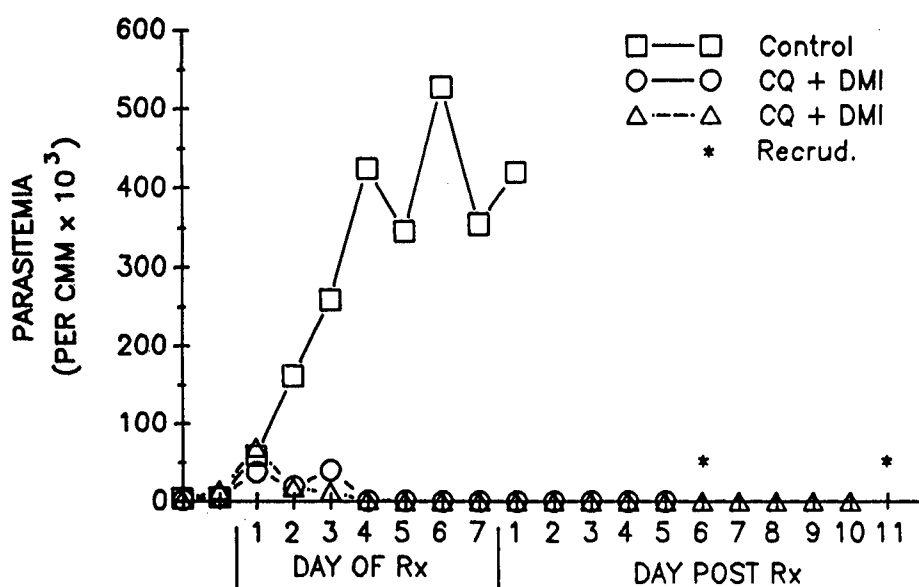
Figure 3. Graphical representation of chloroquine resistant falciparum malaria parasitemia in untreated (control) aotus monkeys and aotus monkeys treated with the combination of chloroquine (CQ) and desipramine (DMI) for seven consecutive days. Note that the drug combination resulted in a rapid clearance of parasitemia, but total cures were not obtained. One monkey recrudesced six days after treatment (Rx) and the second monkey recrudesced 11 days after treatment.

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

The future direction of antimalarial drug development will include basic and applied reserach aimed toward elucidating the modes of action and mechanisms of resistance of known and developmental drugs. These studies will hopefully determine the structure and function of P-glycoprotein-like or other transport proteins in drug resistant parasites. Using molecular modeling and lead-directed synthesis, the design and synthesis of new drugs will result in the development of nontoxic reversal agents or antimalarial drugs which circumvent the malaria parasite's drug efflux mechanism.

Table 5. Comparison of plasma half-lives, bioavailability and fifty percent inhibitory concentrations (IC_{50}) against <u>Plasmodium falciparum</u> of commonly used calcium antagonists and antimalarials.

Compounds	Half-life in Hours	Peak Plasma Concentration (ng/ml)	Parasite IC_{50} (ng/ml)
Resistance Modulator			
Chlorpromazine	30	30-350	1300-1400
Desipramine	18	30-50	1100-1300
Verapamil	2-5	125-40	3300-9200
Cyproheptadine	Unknown	Unknown	200-1000
Ketotifen	Unknown	Unknown	20-500
Diltiazem	3-4	30-200	3400-18000
Nifedipine	2	90-120	15400-35000
Antimalarial Drugs			
Chloroquine	140-280	100-250	3-120
Quinine	9-13	2000-13000	10-120
Tetracycline	6-12	3000	500-18000

MILHOUS, KYLE, MARTIN, GROGL, ODUOLA, ROSSAN & SCHUSTER

REFERENCES

1.   D. Payne, Parasitol. Today 2: 241-246 (1987)
2.   E. F. Boudreau et al., Lancet 2: 1335 (1982).
3.   A. M. J. Oduola et al., Lancet 2: 1304-1305 (1988).
4.   R. E. Desjardins et al., Antimicro. Agents Chemotherap. 710-718 (1979).
5.   W. K. Milhous et al., Antimicrob. Agents Chemother. 27: 525-530 (1985).
6.   W. K. Milhous et al., Prog. Clin. Biol. Res. 61-72 (1989).
7.   A. M. J. Oduola et al., Exper Parasitol. 66: 86-95 (1988).
8.   D. S. Peterson, W. K. Milhous and T. E. Wellems, Proc. Nat. Acad. Sci.(in press, 1990).
9.   T. M. Cosgriff et al., Am. J. Trop. Med. Hyg. 34: 692-693 (1985)
10.   I. Pastan and M. Gottesman, New Eng. J. Med. 316: 1388-1393 (1987).
11.   A. Fojo et al., Cancer Res. 45: 3002-3007 (1985).
12.   S. K. Martin, A. M. J. Oduola and W. K. Milhous, Science 235: 899-901 (1987).
13.   D. E. Kyle, A. M. J. Oduola and W. Milhous, Abstr. FrP-3-13 XIIth Int. Cong. for Trop. Med. and Malaria, Amsterdam, The Netherlands, Sep 18-23, p. 351 (1988).
14.   D. E. Kyle, H. K. Webster and W. K. Milhous, Abstr. No. 518 Amer. Soc. Trop. Med. Hyg., Honolulu, Hawaii, Dec. 10-14, p 322 (1989).
15.   C. M. Wilson et al., Science 244: 1184-1186 (1989).
16.   S. J. Foot et al, Cell 57: 921-930 (1989).
17.   S. J. Foot et al., Proc. Nat. Acad. Sci. (in press, 1990).
18.   D. J. Krogstad et al., Science 238: 1283-1285 (1987).
19.   D. E. Kyle, A. M. J. Oduola, and W. K. Milhous, Trans. Roy. Soc. Trop. Med. Hyg. (in press, 1990).
20.   W. Peters et al., Lancet ii: 334 (1990).
21.   R. N. Rossan et al., Am. J. Trop. Med. Hyg. 34: 1037-1047 (1985).
22.   R. N. Rossan et al., Abstr. No. 359 Am. Soc. Trop. Med. Hyg., Los Angles, Ca, Nov. 29-Dec. 3, p. 236 (1987).
23.   A. J. Bitonti et al., Science 242: 1301-1303 (1988).
24.   M. M. Gottesman and I. Pastan, Biol. Chem. 268: 12163-12166 (1988).

Novel Laser-Based Detector for Gas Chromatography

*Dr. Andrzej W. Miziolek
Dr. Jeffrey B. Morris
Dr. Brad E. Forch
Dr. Randy J. Locke[a]
Ballistic Research Laboratory, SLCBR-IB-I
Aberdeen Proving Ground, MD   21005-5066
a. NAS/NRC Postdoctoral Research Associate

INTRODUCTION: The ability of pulsed lasers to produce microplasmas via laser-induced breakdown has been explored almost since the invention of the laser in the early sixties. The majority of this work has employed infrared (Nd:YAG, CO_2) or visible (ruby) lasers to study the laser-induced breakdown properties of gases[1,2] and solids.[3] One major analytical method which resulted from these studies is that of laser-induced breakdown spectrometry (LIBS).[4,5] In LIBS, an analyte gas or aerosol is converted into a microplasma and the emission spectrum generated is used for elemental analysis. Only recently have uv lasers been employed to generate microplasmas.[6-8] The present article describes the use of uv laser-produced microplasmas for the detection of carbon-containing molecules in the effluent gases of a gas chromatography (GC) capillary column.

A laser-produced microplasma results from the focusing of a high peak-power laser pulse of sufficient energy into a gaseous sample. The microplasma phenomenon is in evidence by the observation of a visible "spark" of broadband light accompanied by an audible "pop" or acoustic report. Laser-produced microplasma formation is a complex process that is generally well-understood and involves the following steps.[1-3] The microplasma is initiated by the presence of free electrons early in the laser pulse. While free electrons may be produced by any means, the usual means of free electron production is via non-resonant multiphoton ionization (MPI) where several photons are absorbed by any atomic or molecular species within the laser focal volume resulting in the ionization of the intact species or its fragments. When a free electron interacts with an ion in such a manner that the electron-ion pair may be regarded as a continuum system, the absorption of a photon results in a transition to a higher continuum state of the system. This process, called inverse bremsstrahlung, results in an increase of the kinetic

energy of the electron. Collisional excitation and ionization occur when the electron acquires enough energy (via inverse bremsstrahlung) to ionize any atomic or molecular species through electron impact; the net effect of collisional ionization is the production of an ion and an additional electron. The cascade, or avalanche, ionization which occurs as the free electrons within the laser focal volume repeat the processes of inverse bremsstrahlung and collisional ionization results in a high degree of ionization within the microplasma. Once the electron population reaches a certain level (ca. 10^{15}/cc), the electrons become the primary absorbers of the remaining laser pulse, heating the focal volume to a very high temperature.[1] Recent work in our laboratory has shown that lasers operating in the ultraviolet can produce resonant microplasmas whose characteristics are different than those produced via non-resonant MPI.[6] In particular, the uv-generated microplasmas require considerably less incident laser energy (ILE) for their production and the energy can be deposited into the microplasma in a more controllable manner than in the non-resonant case. In order to use the uv laser-produced microplasma for the detection of carbon-containing gaseous flows, it is necessary that the microplasma formation occurs only in the presence of a carbon-containing molecule, but not in the case of the carrier gas by itself. This condition is met in regular gas chromatography which uses helium, a high ionization potential gas (IP = 24.5 eV), as the carrier, minimizing the possibility of gas ionization. Since most organic molecules have IPs in the seven to ten electron volt range, gas ionization is much easier to achieve in their presence than in the presence of helium alone.

A uv laser-produced microplasma displays three characteristics which may be exploited for its detection. A large fraction of the gas is ionized, resulting in the production of both ions and electrons.[1] The presence of these charged species can be detected with an optogalvanic probe. The blastwave (acoustic report) associated with the microplasma can be detected by using a photoacoustic detector, such as a microphone. The microplasma also emits electromagnetic radiation which can be detected by photometric means.

Previously, excited and ground electronic state molecular and/or atomic fragments have been observed in the laser photolysis (at 193 and 248 nm) of organic,[9-11] organometallic,[10,12] and organophosphorus[13,14] species. In particular, the ArF laser-induced atomic carbon $^1P_1^0 \rightarrow {}^1S_0$ transition at 248 nm has been used for the photolytic detection of several organic molecules.[11] This transition results from a coincident pumping by the ArF laser of the $^1D_2$ state of carbon atoms to the $^1P_1^0$ state at 193.1 nm (see Fig. 1). This coincidental overlap gives rise to part of the signal at 248 nm that we use for our photometric detection, the remaining being due to plasma atomic emission.
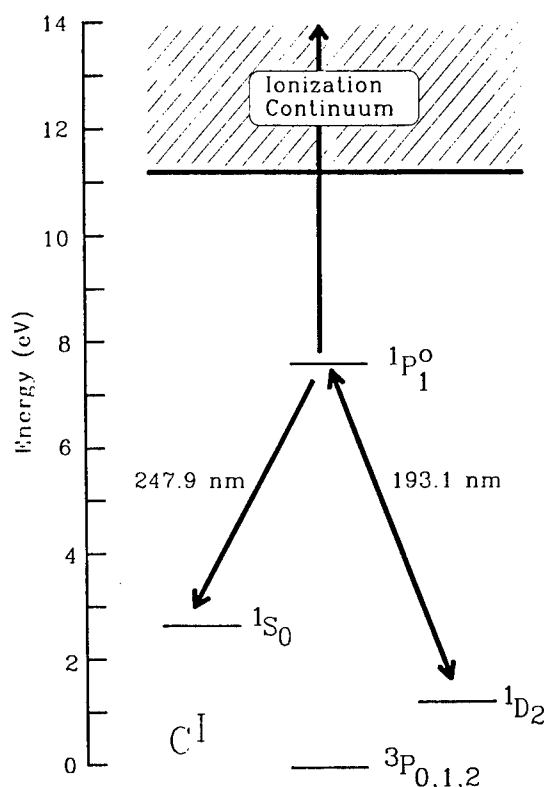
Figure 1. Partial energy level diagram for carbon atoms.

The goal of this paper is to investigate the optogalvanic, photoacoustic, and photometric (248 nm) techniques for the detection of uv laser-produced microplasmas, specifically with the potential development of a new GC microplasma-based detector in mind. For each detection scheme, the linearity of the detector as well as the detection limit for acetylene gas is reported.

EXPERIMENTAL: The experimental apparatus used in this study is detailed below. An ArF excimer laser (Lumonics, model TE 861M-3) supplies the 193 nm pulses (10 pulses per second, 15 nsec duration) used to produce the microplasmas. The laser is equipped with unstable resonator optics to give a collimated beam 1/8" x 3/4" in size. Although the laser is capable of well over 100 mJ pulse energy, for these studies an incident laser pulse energy of approximately only 10 mJ is used. The laser pulse repetition rate of 10 Hz is chosen primarily

for convenience whereas the current limits on repetiton rate for commercially available lasers are approaching the 1 kHz range. Even at 10 Hz, the capillary peaks of approximately 1 second duration are easily resolved and profiled. The laser pulses are focused through a 1" diameter, 50 mm focal length uv grade synthetic fused silica lens (Melles-Griot). The focus of the beam is directed about 0.5 mm above the tip of a 5 meter, 530 micron diameter capillary column (Hewlett-Packard) where the microplasma formation takes place. The laser beam is terminated into a 4" diameter volume absorbing calorimeter (Scientech), which also serves to monitor the laser energy.

A blow-up of region around the tip of the capillary column is shown in Fig. 2. The helium carrier gas (Spectra Gases, 99.999%) flows through the capillary column at a rate of about 10 ml/minute. The column is surrounded by a 3/8" diameter helium sheath (flow rate: 200 ml/minute) to prevent the mixing of room gases (especially O_2) with the effluent gas from the capillary. Without this sheath gas and at these flow rates, the entrainment of room air within the effluent gas from the capillary leads to the condition of continuous breakdown, even during the flow of the helium carrier gas alone. The reason for this is that O_2 is easily ionized by the ArF laser due to the absorption of two photons via the intermediate Schumann-Runge bands. The carrier gas is used without further purification while the sheath gas, which consists of tech-grade, oil-free helium, is sent through a series of filters consisting of activated charcoal, molecular sieves, and dessicant. Centering of the column within the sheath is achieved by the use of a wire screen; the screen allows for unrestricted flow of the helium sheath. The helium carrier and sheath both flow into atmospheric pressure.
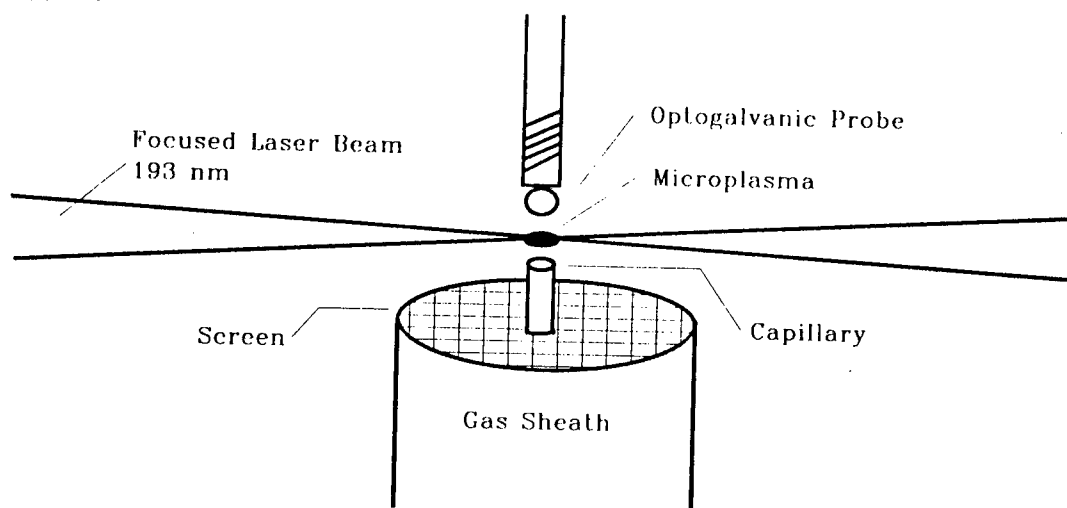

Figure 2. Enlargement of the capillary/laser focal volume
interaction region (capillary diameter = 530 microns).

To determine the microplasma volume, a small continuous flow of acetylene is directed through the capillary column along with the usual helium carrier. The microplasma is observed under magnification using a microscope and the dimensions of the microplasma are estimated by visual comparison with the capillary column. The boundaries of the microplasma are fairly well determined bt the diameter of the capillary column (530 microns), with a slight encroachment of the microplasma into the sheath flow due to diffusion of the acetylene. The diameter of the microplasma is determined by the focused laser beam waist and is estimated to be about 50 microns. The total volume of the microplasma is thus calculated to be approximately 10^{-6} cubic centimeters.

The microplasma is detected via one of three methods. An optogalvanic probe, designed and built by Professor T.A. Cool,[15] is used to detect free electrons produced within the microplasma. The platinum tip (anode) of the probe is biased to +300 volts and is positioned 0.5 mm above the microplasma. The output from the optogalvanic probe is amplified using a differential preamplifier (PAR, model 113) and then signal averaged using a Stanford Research Systems (SRS) boxcar integrator.

For the second detection scheme, a microphone system (B&K Instruments, Inc., type 4133 microphone condenser, type 2615 input stage, and type 2801 power supply) is employed for photoacoustic detection of the microplasma. The microphone is positioned 1 mm above the microplasma (the closest the microphone can be positioned to the microplasma in order not to clip the laser beam). The signal from the microphone is amplified using a Keithley current amplifier (model 427) and is then signal averaged using the SRS boxcar.

Photometric detection of the microplasma is employed by monitoring emission of the atomic carbon $^1P^0_1 \rightarrow {}^1S_0$ transition at 248 nm. A two-lens system (collection efficiency = 1.6%) is used to collect emission from the microplasma and focus it on the entrance slits of a 0.25 m monochromator (McPherson, model 218) equipped with a Hamamatsu 955R photomultiplier tube (PMT). The signal from the PMT is signal averaged using the SRS boxcar integrator without prior amplification.

Quantitative samples were prepared by bleeding the sample gas into an evacuated stainless steel manifold equipped with a barocel pressure sensor. The manifold was also equipped with a septum for removing samples. Signal responses for the three detection techniques were measured by injecting a 1-10 microliter sample of acetylene into the helium carrier going into the capillary column.

RESULTS AND DISCUSSION: Figure 3 shows a response plot for the optogalvanic detection scheme in which the mass of acetylene injected

into the carrier gas flow is plotted versus the signal intensity recorded from the boxcar integrator. The response plot for photoacoustic detection of the acetylene microplasma is very similar to that of the optogalvanic technique. The plot for the photometric detection of acetylene is plotted in Fig. 4. The fall-off in the signal level appears to be due, at least in part, to a drop in the ArF laser intensity over the duration of the experiment. The use of a cryogenic gas processor, not available in our laboratory, would have greatly improved the long term pulse stability. Sensitivities for the three detection schemes are determined from the response plots and are reported in Table 1, along with estimates of potential for improvement for the techniques. One potential advantage of the laser-produced microplasma technique is its sensitivity to non-combustible gases such as CO and CO_2. The relative detector responses for both of these gases were determined to be within a factor of two for that of acetylene measured under similar conditions using photometric detection at 248 nm.
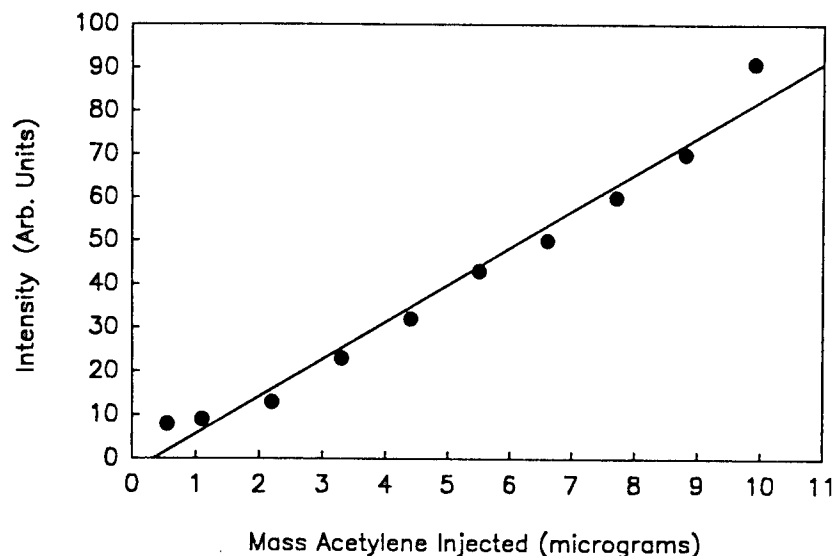

Figure 3. Signal response as a function of injected C_2H_2 mass for optogalvanic detection.

The relatively poor detection limit observed with the photoacoustic detection scheme is caused by a number of factors. Primary among these is the lack of sufficient sensitivity of the microphone used. Also, the longer response time (tens of microseconds) and the acoustic noise generated by the laser contributed to the low S/N situation. Clearly, increasing the ILE would result in a larger plasma with a corresponding

stronger blast wave. Such conditions, however, negate our desire to work with modest laser pulse energies and in addition, the ILE can only be increased up to the point where the carrier gas breaks down.
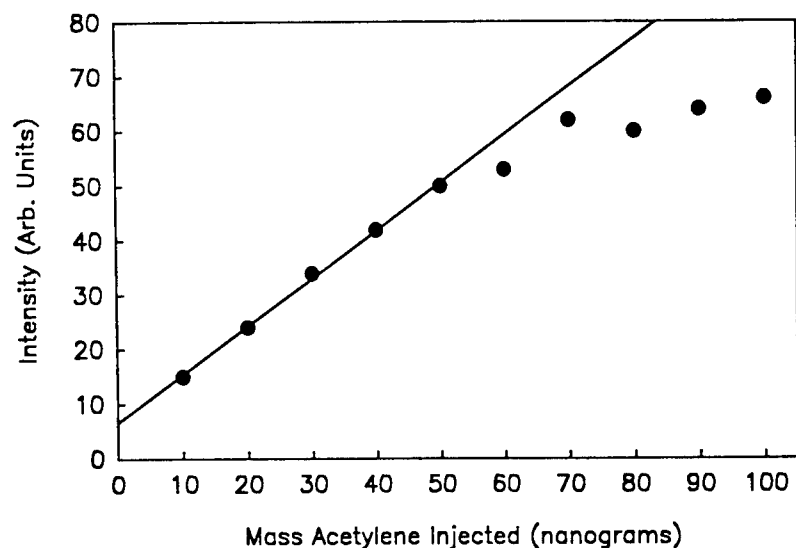

Figure 4. Signal response as a function of injected C_2H_2 mass for photometric detection.

Table 1. Preliminary Detector sensitivities for Optogalvanic, Photoacoustic, and Photometric (248 nm) detection schemes for C_2H_2

Detection Scheme	Sensitivity (ng)	Estimated Potential for Improvement in Sensitivity
Optogalvanic	500	High[a]
Photoacoustic	500	Moderate[b]
Photometric	10	Very High[b]

a. By using different electrode configuration.
b. By using broad-band detection of microplasma radiation and reduction in rf pickup.

The apparent low sensitivity of the optogalvanic detection scheme is somewhat surprising since even in this small volume the electron population within the microplasma is substantial. One explanation for the lack of sensitivity using our optogalvanic detector is the space charge phenomenon in which the electrons in the outer region of the

microplasma repel the core electrons such that most of the electrons are not collected on the small platinum anode bead (1 mm diameter) and are therefore swept away with the carrier gas flow. At the observed detection limit of 500 ng, we estimate that about 10^{11} electrons should be produced (assuming single ionization) from the acetylene molecules which are present within the laser focal volume when the laser fires. The expected peak current from the optogalvanic probe (RC time constant = 20 μ sec) resulting from the collection of 10^{11} electrons would be 0.5 mA. This peak current would correspond to 250 mV after being amplified by a factor of 10 and terminated into 50 Ohms. Taking the sensitivity of the SRS boxcar integrator system to be 5 mV, it is estimated that less than 2% of the electrons produced within the microplasma are being collected. As is indicated in Table 1, this problem can be addressed by using a modified electrode configuration, such as parallel plates, to increase the electron collection efficiency. The sensitivity of the photometric detection technique at 248 nm, while being better than the optogalvanic and photoacoustic techniques by a factor of 50, is still not nearly as high as our estimates indicate it should be. However, our present detection limit, using 10 mJ ILE at 193 nm, indicates a sensitivity similar to LIBS work done on chlorine- and fluorine-bearing molecules which required 100 mJ ILE at 1064 nm.[5] Both the photometric and optogalvanic detection techniques are affected by rf pickup from the excimer laser source. This problem is being addressed by shielding the laser better, particularly the power cord, which has been reported to be a rf noise source for older lasers such as ours.[16]

CONCLUSIONS: We have demonstrated the viability of the concept of uv laser-produced microplasmas as detectors for carbon-containing effluent gases from a gas chromatograph. Optogalvanic, photoacoustic, and photometric means can be used to detect the presence of a microplasma. It appears that photometric detection of laser-produced microplasmas deserves further consideration as a detection method for gas chromatography. A two to four order of magnitude increase in the sensitivity of the photometric detector should be possible by using broad-band detection of the microplasma instead of the wavelength-specific configuration currently utilized. Studies are underway which include optimization of the microplasma optical detection approach, extension of this technique to other hydrocarbons and halogenated molecules, as well as the detection of fragments other than atomic carbon.

REFERENCES

1. C.G. Morgan, Rep. Prog. Phys., Vol. 38, p. 621, 1975.

2. D.C. Smith and R.G. Meyerand, Jr., Princ. Laser Plasmas, George
   Bekefi, Ed., Wiley, New York, p. 457, 1976.

3. P.K. Carroll and E.T. Kennedy, Contemp. Phys., Vol. 22, p. 61,
   1981.

4. L.J. Radziemski, T.R. Loree, D.A. Cremers, and N.M. Hoffman, Anal.
   Chem., Vol. 55, p. 1246, 1983.

5. D.A. Cremers and L.J. Radziemski, Anal. Chem., Vol. 55, p. 1252,
   1983.

6. B.E. Forch and A.W. Miziolek, Comb. Sci. and Tech., Vol. 52,
   p. 151, 1987.

7. A. Biswas, H. Latifi, L.J. Radziemski, and R.L. Armstrong, Applied
   Optics, Vol. 27, p. 2386, 1986.

8. M. Coche, T. Berthoud, P. Mauchien, and P. Camus, Appl. Spectrosc.,
   Vol. 43, p. 646, 1989.

9. J. Schendel and R. Hohmann, E.L. Wehry, Appl. Spectrosc., Vol. 41,
   p. 640, 1987.

10. E.L. Wehry, R. Hohmann, J.K. Gates, L.F. Guilbault, P.M. Johnson,
    J.S. Schendel, and D.A. Radspinner, Applied Optics, Vol. 26,
    p. 3559, 1987.

11. R.C. Sausa, A.J. Alfano, A.W. Miziolek, Applied Optics, Vol. 26,
    p. 3588, 1987.

12. J. Schendel and E.L. Wehry, Anal. Chem., Vol. 60, p. 1759, 1988.

13. S.R. Long, R.C. Sausa, and A.W. Miziolek, Chem. Phys. Lett.,
    Vol. 117, p. 505, 1985.

14. R.C. Sausa, A.W. Miziolek, and S.R. Long, J. Phys. Chem., Vol. 90,
    p. 3994, 1986.

15. P.J.H. Tjossem, T.A. Cool, Chem. Phys. Lett., Vol. 100, p. 479,
    1983.

16. C.N. Merrow, private communication.

Moore and Ledford

Additive Solutions For the 21 Day Preservation of Previously
Frozen Red Blood Cells (U)

*Gerald L. Moore, Dr., and Mary Edith Ledford, Mrs.
Blood Research Division, Letterman Army Institute of Research
Presidio of San Francisco, CA 94129-3600

INTRODUCTION

Supplying blood to battlefield casualties has been a major
problem for the medical services since World War II. Most
blood used by U.S. forces is obtained from blood donor centers
in the Continental United States (CONUS). From the mid 1940's
to 1978 all blood was drawn in acid-citrate-dextrose (ACD) or
citrate-phosphate-dextrose (CPD) and had a 21 day dating
period when stored at refrigeration temperatures. Units of
blood were shipped on ice to CONUS Armed Services Whole Blood
Processing Laboratories (ASWBPL's), where the units were group
and type verified, re-iced, and shipped to the theater of
operations. In theater, blood moved from the depot through
sub-depots to the field hospitals or medical treatment
facilities (MTF's).[1] This involved, multilevel,
transportation system consumed between one and two weeks of
the product's storage life, and left very little storage life
in the end-user's refrigerator. Since combat casualty rates
fluctuate widely, are frequently unanticipated, and the blood
had a very short shelf life, inventory control of blood was
very difficult. Surgical hospitals were frequently cycling
between shortages, and in the meantime discarding numerous
outdated units.
    The dating, or shelf life, of stored blood is determined by
the Food & Drug Administration (FDA) from in vivo studies of
the percentage of stored cells which survive for 24 hours
after reinfusion into the donor. From WW II until 1985 the
FDA set maximum dating as the time when the average unit would
have a percent viability, or post-transfusion survival, of at
least 70%. In 1985 the average was increased to 75%.
    In 1978 the FDA approved a new anticoagulant-preservative
solution which permitted 35 day blood storage. This solution,

Moore and Ledford

CPDA-1, was developed at Letterman Army Institute of Research (LAIR), and helped reduce outdating in military hospitals and civilian blood banks.[2] In 1983 additive solutions were introduced for 35 or 42 day storage of packed red cells. These solutions contain adenine, glucose, sodium chloride, and chemicals to retard lysis (rupture). The solutions were added to packed red cells from blood which had been drawn in CPD and separated into blood components. Unfortunately, the fundamental logistical system for moving blood from the donor in CONUS to the surgical hospital in theater has not changed, and the projected time to move blood in future conflicts remains at 1-2 weeks.

Twenty years ago researchers at the American Red Cross Research Labs and the Naval Blood Research Lab showed that red cells could be stored frozen for long periods in 6 Molar glycerol at -80°C. This method of red cell preservation is approved by the FDA for 10 years of storage, and recent data has indicated that the cells may be cryopreserved for at least 21 years.[3,4] Before these cells can be used, they must be[1] thawed at 37°C and (2) washed free of glycerol by centrifugal or agitational methods. This procedure requires about 30 minutes per step, with the second step being labor intensive. The thawed-washed red cells are in saline/glucose solution, and must be used within 24 hours to meet FDA requirements, since the cells do not have adequate nutrient support, and the sterility "may" have been compromised in the washing procedure.[5]

During development of the DOD Deployable Medical Systems (DEPMEDS) the Military Field Medical Systems Standardization Steering Group requested that the Armed Services Blood Program Office (ASBPO) develop a new plan to deliver blood products in future conflicts. The result was Military Blood Program 2004 (MBP2004), a 20 year planning document which became policy in May 1985. Because of the possible needs for large amounts of blood on short notice, i.e., if a future conflict were to break out in Europe or Asia, the MBP2004 plan calls for large stockpiles of frozen group O red cells at several locations worldwide. These frozen units, already in theater, would be used to meet the blood needs of the military during the initial weeks of fighting, until CONUS sources could be geared-up. Each DEPMEDS is being equipped with one 500 unit -80°C freezer and four Haemonetics Model 115 cell washers. The present concept is to ship the frozen cells from theater depots through the Blood Shipment Units (BSU) to the Medical Treatment Elements (MTE), the new name for tho old MTF's.

Moore and Ledford

Deglycerolization could occur at the BSU's or the MTE's and the cells used at MTE level. Unfortunately, the DEPMEDS have not been staffed with additional personnel to carry out this labor-intensive task. Low-temperature freezers frequently malfunction if moved or used under adverse conditions.

We are working on two modifications of the frozen blood technique which will make its use in the field faster, safer, and more versatile. Our concept is to modify the post-thaw washing procedure to produce a sterile product in a faster, less labor-intensive manner, and to develop an additive solution for restorage of the thawed-washed red cells which will allow a dating period of 3 weeks at 4°C. These changes would enable frozen blood to be processed at the more stable theater-depot level, then shipped to DEPMEDS or other end-users as liquid blood with a 3 week dating. This would move the biotechnology out of the mobile medical treatment facilities so that they could concentrate on treating patients, with the assurance of a safe, reliable supply of blood.

The development of membrane based washing, an approach which may eventually replace centrifugal washing, is ongoing at the Naval Blood Research Laboratory Boston, MA, and at our laboratory, and will be published in detail later. In this report we will focus on the post-thaw additive solutions, and present an overview of some of the significant developments in this research effort.

METHODS AND MATERIALS

Whole units (450 ml) of blood were drawn from healthy adult volunteers with informed consent. The blood was collected into CPDA-1 oversized bag units made by Fenwal (Deerfield, IL) or Cutter Labs (Berkeley, CA). All units were frozen using the Valeri method on the day of donation.[3] Some additional frozen units, which had been frozen by the Meryman procedure, were obtained from the Frank Camp Memorial Blood Center.[6] All units of frozen cells were thawed at 37°C and deglycerolized on a Haemonetics Model 115 cell washer, using the appropriate techniques.[3,6] The washed cells had a hematocrit of 45-50% and were repacked to Hct=75 by centrifugation for 10 min at 3500 g. Additive solutions were introduced by sterile filtration. In some studies, units of packed cells were divided into halves or thirds for comparative treatments before introducing the additive

Moore and Ledford

solution. The volume of additive solution was 50-100 ml per
unit of red cells. All red cell units were assayed after
mixing with the additive solution, and weekly for 3 weeks,
with storage at 4°C between sampling.

Additive solutions consisted of ADSOL, as purchased from
Fenwal Laboratory (100ml; containing 2.2g glucose, 0.75g
mannitol, 27mg adenine and 0.9g sodium chloride), or an
additive solution (AS) made by us with varying concentrations
of adenine, glucose, sodium phosphate, sodium chloride,
mannitol, citric acid and/or sodium citrate.

Analyses of red cell adenosine triphosphate (ATP),
2,3-diphosphoglycerate (2,3-DPG) and glucose were done by
fluorometric enzyme assays.[7] Supernatant hemoglobin was
measured by a micro cyanmet method,[8] which was used to
determine the percent of red cells which lysed (ruptured).
Red cell morphology was measured by the method of Hogman.[9]
Osmolarity was measured on a Model 3D II osmometer (Advanced
Instruments Inc., Needham, MA) and pH was measured at 37°C in
a Corning Blood Gas Analyzer. At the end of 4°C storage, each
bag of red cells was cultured to assure sterility.

The AS formulation was optimized using the experimental
design generating program, COED (CompuServe, Corp., Columbus,
OH). This program uses the statistical principles of
D-optimality to generate modified fractional factorial
designs. The program considers linearity, nonlinearity, and
interaction of all variables and produces a limited subset of
all possible experiments to define the design space at any
desired probability level (we used 0.90). Experimental
results from the COED generated experiments were analyzed by
multilinear regression techniques (BMDP Co. Los Angeles, CA)
and the resulting equations were optimized mathematically.[10]
This approach has been defined elsewhere.[7] Other statistical
analysis of data was performed using BMDP 2V repeated measures
analysis of variance, with multiple comparisons done by
Fisher's least significant difference test. If variances over
the repeated measure (time) were significantly different,
comparisons were made using the Student's t-test. Alpha (Type
I error) for all tests was set at 0.05.

RESULTS AND DISCUSSION

Fresh-frozen red cells which have been thawed and washed
have intracellular ATP and 2,3-DPG levels similar to that of
fresh blood. ATP has been shown to correlate with red cell
viability, in the sense that higher ATP levels usually

Moore and Ledford

indicate higher post-transfusion survival of the cells.
However, there are exceptions to this relationship, and ATP
levels below 30% of initial usually indicate cells with
viabilities below 70%. The cellular 2,3-DPG level has a very
high correlation with the ability of the cells to deliver
oxygen to tissues. The loss of all 2,3-DPG represents about a
50% reduction in the ability of the red cell to deliver oxygen
at tissue partial pressures (all other things being equal).
Thus ATP and 2,3-DPG serve as in vitro correlates of red cell
viability and function respectively.

In our initial studies (Fig. 1) the levels of ATP and
2,3-DPG are shown for red cells stored in the final
saline/glucose wash solution (CTR), ADSOL, and an initial
formulation of AS. ATP was well preserved in AS, but less so
in ADSOL, and was inadequate in the control after 7 days. The
use of ADSOL to store fresh, or never-frozen, blood produces
better maintenance of ATP, implying that the freeze-thaw
process has stressed or damaged the metabolic mechanisms of
the red cells. The 2,3-DPG levels drop to less than 20% by 7
days in both the control and the ADSOL treated red cells.
This drop is also seen in never-frozen red cells stored in
ADSOL after 2 weeks of storage.[11] The 2,3-DPG levels of red
cells in the AS system were significantly better at 7 and 14
days.

Freshly drawn CPD or CPDA-1 blood has a pH of 7.0 to 7.1 at
37°C. Fresh-frozen/thawed/washed red cells have a pH of 6.6.
Obviously the extensive washing (2 liters of saline per 200 ml
of cells) removes not only glycerol, but also any residual
plasma, which helps buffer the cells, and much of the internal
metabolites such as glucose, pyruvate, and other small
nutrient species. This results in a cell that is
metabolically deprived and too acidic. Such cells are more
susceptible to lysis than fresh nonfrozen red cells.
Supernatant hemoglobin levels are shown in Fig. 2 for cells
stored in saline/glucose (CTR), ADSOL or AS. The use of
mannitol in the ADSOL and AS systems does reduce lysis
compared to CTR, but the values still exceed the 1% levels
(about 300mg/dl) currently considered maximal by the FDA.

We have done extensive studies to optimize the AS
formulation. The COED generated designs were run using
several levels of the component species. Regression equations
were developed for the ATP, 2,3-DPG, and %lysis levels
generated by the varying component concentrations. Since
these output variables work against each other (i.e.,
improving ATP generally reduces 2,3-DPG), each output variable

POST-THAW PRESERVATION
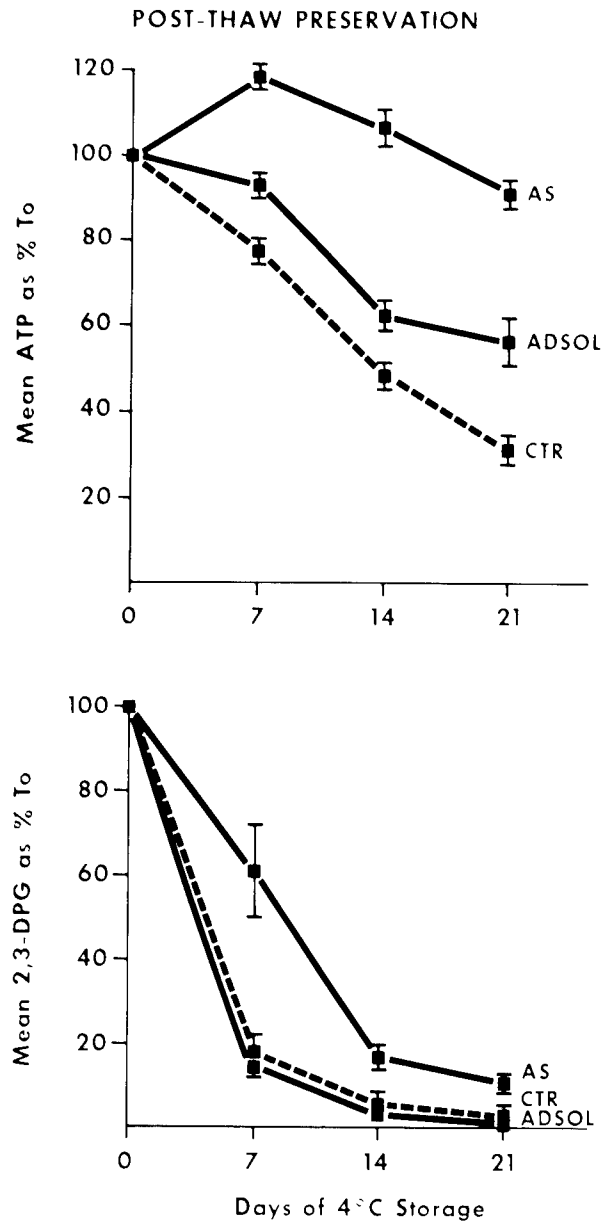
1. Red cell ATP and 2,3-DPG during 21 days of post-thaw preservation at 4°C. Cells in one of three solutions: control (CTR) was isotonic saline with 200 mg/dl glucose, Fenwal's ADSOL solution, and an additive solution (AS) containing adenine, glucose, tribasic sodium phosphate, mannitol and sodium chloride.
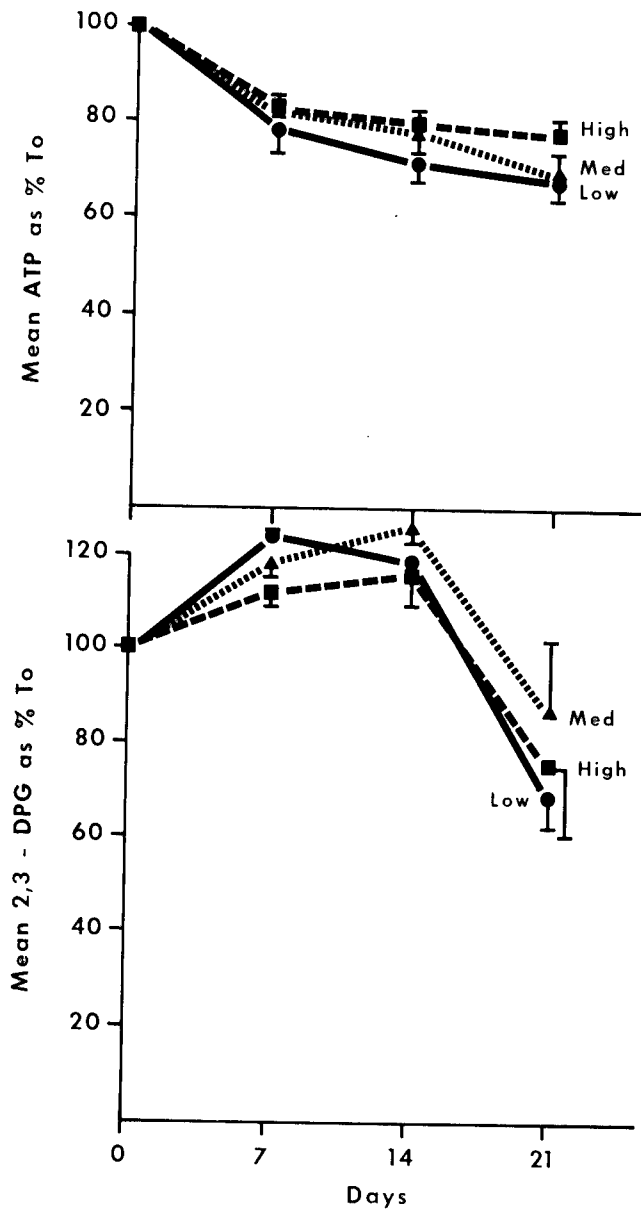
Moore and Ledford

POST-THAW PRESERVATION
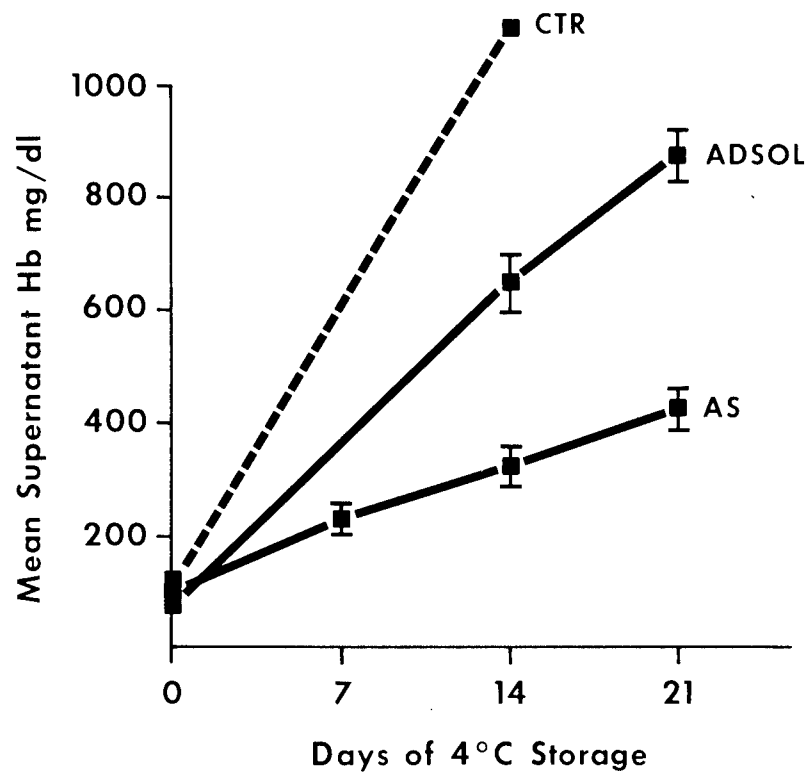
2. Supernatant hemoglobin during post-thaw of red cells in control, ADSOL or AS solutions. A 1% lysis of red cells in this study was equivalent to about 300 mg/dl of hemoglobin.

Moore and Ledford

3. Red cell ATP and 2,3 DPG during post-thaw storage in a mathematically optimized additive solution (Med) and two bracketing solutions (Low and High).
are shown in Table 1.

Moore and Ledford

Table 1

Mean ± SEM Values for Red Cells in Two AS Systems

	Day-0	Day-21 (in AS)	Day-21 (in AS+Cit)
% Initial ATP	100	80 ± 6	80 ± 5
%Initial 2,3-DPG	100	85 ± 5	65 ± 3
% RBC Lysis	0	1.5 ± 0.5	0.5 ± 0.2
% Morphology Score	100	82 ± 5	75 ± 4
pH @ 37°C	7.15	6.6 ± .02	6.5 ± .03
Super. K meq/l	3.5	52 ± 6	57 ± 6

was optimized with regard to minimally acceptable levels of
the other two, and a balance obtained to yield the best
combination of ATP and 2,3-DPG retention (data not shown).
The resulting optimal formulation (Med {i.e., medium}) was
tested in the lab, with two formulations "bracketing" the
calculated best formula (Low, High). Results are shown for
ATP and 2,3-DPG responses in Fig. 3. Other parameters of red
cell quality, for the optimal AS formulation, at days 0 and 21
   The optimal formulation for AS produced good maintenance of
both ATP and 2,3-DPG. The ATP support comes from a balance of
adenine, glucose, and phosphate ions. The good 2,3-DPG
maintenance is due to the mixture of metabolites and the
ability of the trisodium phosphate to buffer the excessive
acid generated during the deglycerolizing step. The AS
solution has a pH of 11.3, but when mixed with the washed
cells, the final suspension pH is 7.1 (Table 1). The
retention of ATP in these cells is similar to the profiles of
nonfrozen cells (i.e., in ADSOL for 35 days) and implies that
they will have good viability. The excellent retention of
2,3-DPG in this system is unique and not duplicated in any
existing liquid storage system. Thus, these cells have better
oxygen delivery during their second and third weeks of storage
than nonfrozen red cells, where 2,3-DPG levels decline to
below 20% by 7 days. The statistical correlation between P50

Moore and Ledford

and 2,3-DPG in these studies is 0.90, N=117.

The freeze-thaw process is traumatic to red cell membranes, increasing their tendency to lyse. Mannitol, which is very effective at retarding lysis in fresh red cells, is less effective after a freeze-thaw cycle. Under the most favorable conditions we found a mean lysis rate of 1.5% with large variation between donors. In an effort to further reduce this hemolysis, we studied many membrane stabilizing compounds, including autologous plasma, as additional components of the AS mixture. Plasma, and in particular, a low molecular weight fraction, were shown to be very effective. Further studies suggested that the citrate was responsible for reduced lysis. Substitution of citric acid/Na citrate for mannitol in the AS solution produced a 50% reduction in lysis, no change in ATP and only a 10-20% reduction in 2,3-DPG levels during 21 days of red cell storage.

The in vitro properties of this additive solution appear very promising when compared to similar properties for nonfrozen red cell systems. Therefore, we are starting clinical trials to determine the percent viability of cells stored in this solution for 14 and 21 days. If the red cells have mean viabilities above 75% in our survival studies, and if the question of washing methodology sterility can be resolved, then many new possibilities are available for using frozen red cells.

Frozen cells which are stockpiled in theater depots could be thawed, washed, and repackaged at depot for 21 day usage by MTE's. Equipment maintenance support and product quality control support for this operation could be provided much more easily at depot than in the MTE's. Personnel requirements for processing would be removed from the MTE's and placed at depot where more centralization is possible, and the level of experience much higher. Finally, much greater inventory flexibility would be available to the medical units. This system could also provide an emergency blood supply to the country in times of natural emergencies or civilian shortages.

References

1. Neel S. Medical Support of the U S Army in Vietnam,
Department of the Army, Washington D.C. 1973.

2. Peck C.C., Moore G.L., Bolin R.B. CRC Crit Rev Clin Lab Sci
1981; 13:173-212.

3. Valeri C.R. Methods in Hematology 1988; 17:277-304.

4. Valeri C.R., Pivacek L.E., Gray A.D. Transfusion 1989;
29:429-37.

5. Valeri C.R. Blood Banking and the Use of Frozen Blood
Products. CRC Press 1976.

6. Meryman H.T. Prog. in Hematol. 1979; 11:193-227.

7. Moore G.L., Ledford M.E. Transfusion 1985; 25:319-24.

8. Moore G.L., Ledford M.E., Merydith A. Biochem Med 1981;
26:167-73.

9. Hogman C.F., deVerdier C., Ericson A., Hedlund K. Vox Sang
1985; 48:257-68.

10. Moore G.L. The Polynomial Optimizer, Assess 1983; Jan
pp37-52.

11. Moore G.L. CRC Crit Rev in Clin Lab Sci 1987; 25:211-229.

A Modified Lagrange Pressure Gradient for
the Regenerative Liquid Propellant Gun (U)

*Walter F. Morrison, Dr.
Terence P. Coffee, Dr.
Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

INTRODUCTION

Over the past several years, a number of computer models of the regenerative interior ballistic process have been developed. The models used most frequently at the Ballistic Research Laboratory are those developed by Coffee[1] and Gough.[2] The Coffee model utilizes lumped parameter descriptions for the three main regions in the regenerative gun, the propellant reservoir, the combustion chamber and the barrel. In contrast, Gough has chosen to treat the reservoir and chamber as lumped parameter regions, while utilizing a one-dimensional flow model for the barrel region. In order to model the barrel as a lumped parameter region, a pressure gradient model, analogous to the standard Lagrange gradient used in many solid propellant gun models,[3-4] is required. The pressure gradient model simulates the drop in pressure from the breech or barrel entrance to the projectile base, which results from the velocity gradient developed as the projectile accelerates down the barrel. The development of such a pressure gradient model requires that the unique features of the regenerative interior ballistic (IB) process be incorporated into the model. These features are the non-zero gas velocity at the entrance to the barrel, and the existence of a rarefaction wave traveling along the barrel toward the projectile base after all-burnt.

In the following paper, the development of a modified Lagrange gradient model, which accounts for the unique features of the regenerative IB process, is described. The modified Lagrange gradient has been incorporated into the lumped parameter interior ballistic model developed by Coffee,[1] and computer simulations of a hypothetical 120-mm tank cannon have been conducted. Simulations were also conducted using the Coffee[1] model with a standard Lagrange gradient and the Gough[2] model with a one-dimensional barrel flow model. Using the results from the Gough model as a baseline, comparisons of muzzle velocities, pressure versus time, and detailed velocity and pressure profiles in the barrel are presented and discussed.

DESCRIPTION OF THE RLPG INTERIOR BALLISTIC PROCESS

The RLPG IB process is based on the injection of the propellant into the combustion chamber during the IB cycle. Several RLPG designs involving different piston configurations have been developed and tested. The configuration known as Concept VI is depicted in Figure 1. The system consists of 3 regions: (1) The liquid reservoir bounded by the regenerative piston, the fixed bolt, and the transducer block; (2) the combustion chamber; and (3) the barrel. In regenerative gun concepts similar to Concept VI, there is a large area change (about a factor of 4) from the combustion chamber into the barrel.

Initially, the piston is seated on the front of the bolt, sealing the propellant in the reservoir, see Figure 1. The process is initiated by firing an igniter into the combustion chamber, which generates an initial pressure in the chamber, see Figure 2.

The increasing chamber pressure acts on the regenerative piston, forcing it to the rear, thus pressurizing the propellant reservoir. Due to the area difference across the piston face from the chamber to the reservoir, the piston acts as a pump, resulting in (1) a pressure in the liquid reservoir which is higher than that in the combustion chamber, and (2) injection of propellant into the combustion chamber.

As the piston moves to the rear over the contoured, fixed bolt, the injection area first increases to a maximum value associated with the maximum design pressure of the system, and then decreases, decelerating the piston as it completes its stroke. When the shot-start pressure is exceeded, the projectile begins accelerating down the tube. Analyses[5] of experimental data indicate that early in the ballistic cycle, significant amounts of unburnt propellant may accumulate in the combustion chamber, that this accumulated propellant is rapidly consumed as the pressure rises to
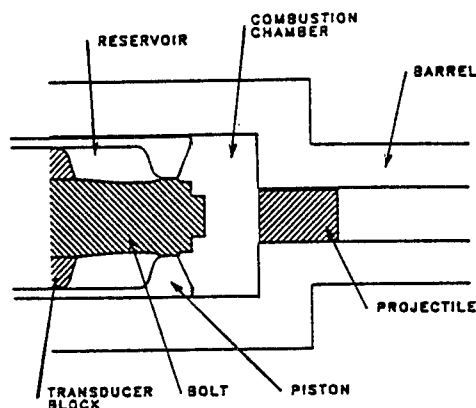

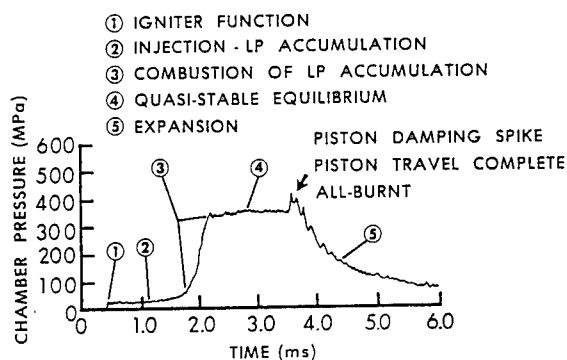Figure 1. Regenerative Liquid Propellant Gun, Concept VI

Figure 2. The Regenerative Interior Ballistic Process

its maximum value, and that propellant subsequently injected is rapidly consumed leading to very low values of propellant accumulation in the latter stages of the ballistic cycle, see Figure 2. Such analyses would also indicate that over <u>most</u> of the projectile travel, propellant is consumed rapidly in the combustion chamber with little accumulation, and, therefore, only small amounts of unburnt propellant would be transferred into the barrel.

Utilizing such analyses, most regenerative IB models have been based on the following assumptions:

1. The combustion chamber is a homogeneous region containing either a two-phase mixture of propellant droplets and combustion gases (if a finite rate droplet burning model is used), or combustion gases only (if an instantaneous propellant burning model is assumed).
2. Only combustion gases enter the barrel region.
3. As the gases enter the barrel, they undergo an expansion process.

The lumped parameter model developed by Coffee does permit a two-phase mixture and droplet combustion in the barrel region, i.e. assumption 2 is relaxed, but this option has not been exercised extensively. The fully one-dimensional IB model recently developed by Gough[6] treats the combustion chamber as a non-homogeneous region, and also permits a two-phase mixture and droplet combustion in the barrel region. Investigations using this model of the effects of relaxing assumptions 1 and 2 are underway, but results are not yet available.

Following the completion of piston motion, the remaining propellant in the combustion chamber (and barrel) is quickly consumed, leading to the "all-burnt" condition. Prior to all-burnt, the gases required to maintain the operating pressure, under the conditions of rapid expansion resulting from piston motion in the chamber and projectile motion in the barrel, are supplied by propellant combustion, primarily in the chamber region. Upon burnout, the combustion chamber pressure rapidly decreases and a rarefaction wave would be expected to move along the barrel toward the projectile base.

This phenomena was originally suggested by Morrison et al,[7] see Figure 3, based on simulations using the one-dimensional model developed by Gough.[2] (It appears similar to the process observed in solid propellant systems utilizing stick propellant charges.) Recent analyses of data from 25-mm regenerative test firings conducted in the mid-1970's provide experimental verification of the existence of such a rarefaction wave,[8] see Figure 4. In these tests, pressure gages were located at several points along the barrel. The resulting pressure profiles are non-monotonic along the barrel, with the point of maximum pressure moving toward the base of the projectile, indicating the presence of a rarefaction wave.
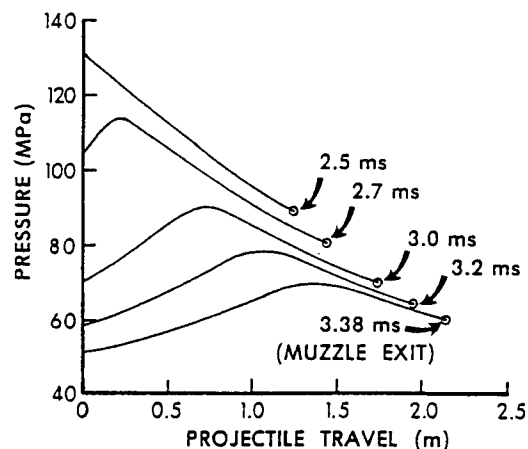

Figure 3. Pressure Profiles from a Simulation of a 25mm RLPG[7]

The ballistic process is completed with muzzle exit and "blow-down" of the gases remaining in the barrel.

MODIFIED LAGRANGE PRESSURE GRADIENT

The standard Lagrange pressure gradient does not accurately reflect the physical process in the barrel region of a regenerative gun. The shortcomings of the standard Lagrange model are as follows:

1. The expansion of the combustion gases from the chamber into the barrel with a large area reduction is not treated.

2. The non-zero gas velocity at the entrance to the tube, and the time dependence of the mass of combustion gases in the barrel are not accounted for.

Figure 4. Pressure Profiles from Experimental 25mm RLPG Firings[8]

3. The rarefaction wave in the barrel after all-burnt is not simulated.

In order to simulate these processes in the interior ballistics model, modifications to the standard Lagrange pressure gradient model are required. In the following development, it is assumed for simplicity that only combustion gases enter the barrel. However, the development applies equally to a two-phase mixture of combustion gases and propellant droplets if the assumption of uniform mixture density is made.

MODIFIED LAGRANGE PRESSURE GRADIENT; BEFORE ALL-BURNT

The equations of motion governing the flow of gas in a gun tube between the barrel entrance and the projectile base are

$$\frac{\partial \rho}{\partial t} + v\frac{\partial \rho}{\partial x} + \rho\frac{\partial v}{\partial x} = 0 \qquad (1) \qquad\qquad \rho\left[\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x}\right] = -\frac{\partial P}{\partial x}. \qquad (2)$$

We must now solve the equations of motion for the gases in the barrel, Equations (1) and (2), with the boundary conditions,

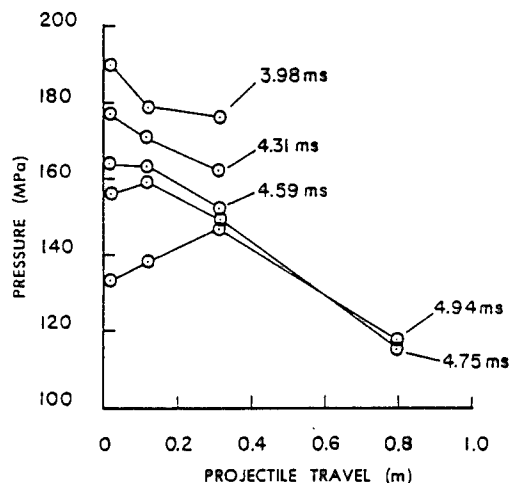$$x_L = 0 \qquad (3) \qquad\qquad x_R = y \qquad (4)$$

$$v_L \neq 0 \qquad (5) \qquad\qquad v_R = u_p. \qquad (6)$$

580

We also utilize the Lagrange assumption, $\frac{\partial \rho}{\partial x} = 0$, i.e. the gas density in the barrel varies with time but not with position along the barrel.[3-4] From the continuity equation, we obtain

$$\frac{\partial v}{\partial x} = -\frac{1}{\rho} \frac{\partial \rho}{\partial t}. \tag{7}$$

Noting that the mass of combustion gases in the barrel, m, is a function of time and that $\rho = m / A_B x_R$, we obtain

$$\frac{\partial v}{\partial x} = -\frac{1}{\rho} \frac{\partial \rho}{\partial t} = -\frac{\dot{m}}{m} + \frac{u_P}{x_R}, \tag{8}$$

where the time rate of change of the mass of the gases in the barrel, $\dot{m}$, is $\rho A_B v_L$. Therefore,

$$\frac{\partial v}{\partial x} = \frac{u_P}{x_R} - \frac{v_L}{x_R}. \tag{9}$$

Integrating on $[0,x]$, we have

$$v(x) = u_P \left(\frac{x}{x_R} \right) + v_L \left(1 - \frac{x}{x_R} \right), \tag{10}$$

and

$$\dot{v}(x) = \dot{u}_P \left(\frac{x}{x_R} \right) + \dot{v}_L \left(1 - \frac{x}{x_R} \right) - u_P(u_P - v_L) \left(\frac{x}{x_R^2} \right) . \tag{11}$$

Substituting Equation (10) and (11) in Equation (2), and noting that

$$\dot{u}_P = (P_{BASE} - P_{RES}) \frac{A_B}{M_P}$$

we obtain,

$$-\frac{1}{\rho} \frac{\partial P}{\partial x} = \left\{ \dot{v}_L \left(1 - \frac{x}{x_R} \right) + \frac{A_B}{M_P} (P_{BASE} - P_{RES}) \left(\frac{x}{x_R} \right) \right.$$

$$- u_P(u_P - v_L) \left(\frac{x}{x_R^2} \right)$$

$$+ \left[v_L + (u_P - v_L) \left(\frac{x}{x_R} \right) \right] \frac{u_P - v_L}{x_R} \right\} . \tag{12}$$

Integrating on $[0,x]$, we have

$$P(x) = P_L - \rho \left\{ \frac{A_B}{2M_p}(P_{BASE} - P_{RES})\frac{x^2}{x_R} \right.$$
$$\left. + \left[\dot{v}_L + \frac{1}{x_R}v_L(u_p - v_L) \right]\left[x - \frac{1}{2}\frac{x^2}{x_R} \right] \right\} \quad . \tag{13}$$

Defining the space mean pressure, $\bar{P}$, by

$$\bar{P} = \frac{1}{y}\int_0^y P(x)dx, \tag{14}$$

and integrating Equation (13) on $[0,x_R]$ we obtain,

$$\bar{P} = P_L - \frac{m}{6M_p}(P_{BASE} - P_{RES}) \quad - \frac{m}{3A_B}\left[\dot{v}_L + \frac{v_L(u_p - v_L)}{x_R} \right]. \tag{15}$$

Similarly, using the definition for the kinetic energy of the gas,

$$KE_{GAS} = \int_0^y \frac{1}{2}A_B\rho v^2 dx, \tag{16}$$

and substituting from Equation (10), we have

$$KE_{GAS} = \frac{m}{6}(u_p^2 + u_p v_L + v_L^2). \tag{17}$$

MODIFIED LAGRANGE PRESSURE GRADIENT; AFTER ALL-BURNT:

In the description of the RLPG interior ballistic process, we discussed the rarefaction wave which travels along the barrel from the chamber toward the projectile base after propellant burnout. We model the rarefaction wave as a discontinuity in the spatial derivatives of velocity and pressure, which travels along the barrel at a velocity equal to the average speed of sound in the gas plus the local gas velocity, see Figure 5. Since this discontinuity resembles a "hinge" in the velocity and pressure distributions, we have labeled it the "hinge point" and quantities at the hinge point are denoted by the subscript H. This representation of a contact discontinuity in a gun tube was recently suggested by Gough.[9] The velocity of the hinge point, $\dot{x}_H$, is then

$$\frac{\partial x_H}{\partial t} = v_H + a. \qquad (18)$$

The hinge point divides the barrel into two region,

Region I: $\quad 0 \leq x \leq x_H$

and

Region II: $\quad x_H < x \leq x_R$.

a. <u>Solution in Region I:</u> We assume that the velocity distribution is linear in both regions I and II, i.e. we use the second form of the "Lagrange Approximation" noted by Corner[3]. From Figure 5, we see that
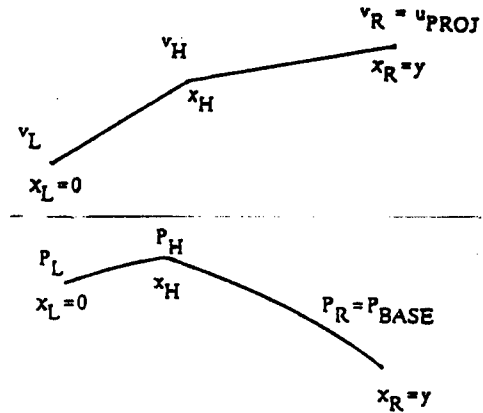

Figure 5. Velocity and Pressure Distributions About Hinge Point

$$\frac{\partial v_I(x)}{\partial x} = \frac{v_H - v_L}{x_H}, \qquad 0 \leq x < x_H. \qquad (19)$$

and integrating on $[0, x_H]$ we obtain

$$v_I(x) = v_H\left(\frac{x}{x_H}\right) + v_L\left(1 - \frac{x}{x_H}\right). \qquad (20)$$

Using (20) in Equation (2), we have

$$-\frac{1}{\rho}\frac{\partial P_I}{\partial x} = \left[\dot{v}_H - \frac{a(v_H - v_L)}{x_H}\right]\left(\frac{x}{x_H}\right) + \left[\dot{v}_L + \frac{v_L(v_H - v_L)}{x_H}\right]\left(1 - \frac{x}{x_H}\right). \qquad (21)$$

We have assumed that mass entering the barrel after all-burnt is distributed uniformly over the barrel region, such that there is a uniform gas density in the barrel, consistent with the basic Lagrange assumption. Integrating (21) on $[0,x]$ gives us the pressure distribution in Region I,

$$P_I(x) = P_L - \rho\left\{\left[\dot{v}_H - \frac{a(v_H - v_L)}{x_H}\right]\left(\frac{x^2}{2x_H}\right) + \left[\dot{v}_L + \frac{v_L(v_H - v_L)}{x_H}\right]\left(x - \frac{x^2}{2x_H}\right)\right\}, \qquad (22)$$

where

$$\dot{v}_H \doteq \frac{\partial}{\partial t}\left\{v_{II}(x)\Big|_{x_H}\right\}. \qquad (23)$$

583

  b. <u>Solution in Region II</u>: It is assumed that at burnout $x_H = x_L = 0$, and that the hinge point, x_H, then propagates down the barrel toward the projectile base at a velocity given by Equation (18). Since x_H propagates at the speed of sound plus the local gas velocity, Region II is acoustically isolated from events in the combustion chamber. Therefore, we assume that the linear velocity gradient in Region II is fixed by the value of v_L at the time of burnout, v_{Lb}, such that

$$\frac{\partial v_{II}(x)}{\partial x} = \frac{v_R - v_{Lb}}{x_R}, \qquad x_H \le x \le x_R. \tag{24}$$

Then,

$$v_{II}(x) = v_R\left(\frac{x}{x_R}\right) + v_{Lb}\left(1 - \frac{x}{x_R}\right), \tag{25}$$

$$v_H = v_R\left(\frac{x_H}{x_R}\right) + v_{Lb}\left(1 - \frac{x_H}{x_R}\right), \tag{26}$$

and

$$\dot{v}_H = \left(\frac{v_R - v_{Lb}}{x_R}\right)(v_H + a) + \dot{v}_R\frac{x_H}{x_R} - v_R(v_R - v_{Lb})\frac{x_H}{x_R^2}. \tag{27}$$

We now use (25) in Equation (2) to obtain,

$$-\frac{1}{\rho}\frac{\partial P_{II}}{\partial x} = \dot{v}_r\left(\frac{x}{x_R}\right) + \left[\dot{v}_{Lb} + \frac{v_{Lb}(v_R - v_{Lb})}{x_R}\right]\left(1 - \frac{x}{x_R}\right) \tag{28}$$

and

$$P_{II}(x) = P_H - \rho\left\{\dot{v}_R\left(\frac{x^2 - x_H^2}{2x_R}\right) + \left[\dot{v}_{Lb} + \frac{v_{Lb}(v_R - v_{Lb})}{x_R}\right]\left[(x - x_H) - \frac{x^2 - x_H^2}{2x_R}\right]\right\} \tag{29}$$

where,

$$v_R = u_p \tag{30}$$

$$\dot{v}_R = \dot{u}_p = \frac{(P_{BASE} - P_{RES})A_B}{M_p}, \tag{31}$$

and from (22), we have

$$P_H = P_L - \rho\left\{(\dot{v}_L + \dot{v}_H)\frac{x_H}{2} + \frac{1}{2}(v_H - v_L)(v_L - a)\right\}. \tag{32}$$

c. <u>Space Mean Pressure</u>: We again use the definition of Space Mean Pressure, Equation (14),

$$\overline{P} = \frac{1}{x_R} \int_0^{x_R} P(x)dx = \frac{1}{x_R} \left\{ \int_0^{x_H} P_I(x)dx + \int_{x_H}^{x_R} P_{II}(x)dx \right\} \ . \tag{33}$$

Substituting from Equations (22) and (29) we have,

$$\overline{P} = P_L - \frac{\rho x_H}{2} \left\{ \left[\dot{v}_H - \frac{a(v_H - v_L)}{x_H} \right] \left(1 - \frac{2x_H}{3x_R} \right) + \left[\dot{v}_L + \frac{v_L(v_H - v_L)}{x_H} \right] \left(1 - \frac{x_H}{3x_R} \right) \right\}$$

$$- \frac{\rho x_R}{6} \left\{ \dot{v}_R \left[\left(1 - \frac{x_H}{x_R} \right)^2 \left(1 + \frac{2x_H}{x_R} \right) \right] \right.$$

$$\left. + 2 \left[\dot{v}_{Lb} + \frac{v_{Lb}(v_R - v_{Lb})}{x_R} \right] \left(1 - \frac{x_H}{x_R} \right)^3 \right\}. \tag{34}$$

d. <u>Kinetic Energy of the Gas in the barrel</u>: The Kinetic Energy of the gas is given by Equation (16),

$$KE_{GAS} = \int_0^{x_R} \frac{1}{2} A_B \rho v^2 dx = \frac{1}{2} A_B \rho \left\{ \int_0^{x_H} v_I^2(x)dx + \int_{x_H}^{x_R} v_{II}^2(x)dx \right\}. \tag{35}$$

Substituting from (20) and (25), we have,

$$KE_{GAS} = \frac{m}{6} \left\{ [v_H^2 + v_H v_L + v_L^2] \left(\frac{x_H}{x_R} \right) \right.$$

$$+ v_R^2 \left[1 - \left(\frac{x_H}{x_R} \right)^3 \right] + v_R v_{Lb} \left[1 - 3 \left(\frac{x_H}{x_R} \right)^2 \left(1 - \frac{2x_H}{3x_R} \right) \right]$$

$$\left. + v_{Lb}^2 \left(1 - \frac{x_H}{x_R} \right)^3 \right\}. \tag{36}$$

DESCRIPTION OF TEST CASE

In order to assess the modified Lagrange gradient model, a series of test cases was developed. Since the largest effect would be expected in high velocity, high charge-to-mass ratio situations, a hypothetical 120-mm regenerative LP tank cannon used in an earlier study[10] was selected. The charge mass and other gun parameters, with the exception of liquid injection area, were held constant. Three projectile masses were used to provide a range of velocities, and the liquid injection area was adjusted for each projectile mass to produce a maximum chamber pressure of 500 MPa. A hypothetical liquid gun propellant was also used in the simulation. This propellant has the thermochemistry of JA-2 and the physical properties of a HAN-based

LP. This choice was made to provide somewhat higher projectile velocities than would be achievable with a HAN-based LP in the specific cannon configuration used in the study. The gun parameters and thermochemical data used in the simulations are provided in Tables 1 and 2.

TABLE 1. 120-mm Cannon Parameters

Caliber	120 mm
Propellant Mass	16.8 kg
Projectile Mass*	5, 7, 13 kg
Projectile Travel	6.3 m
Initial Chamber Vol	6000 cm³
Reservoir Volume	11700 cm³
Chamber Diameter	34.2 cm
LP Injection Area*	94, 76, 60 cm²
Max Chamber Pressure	500 MPa
Max Reservoir Pressure	700 MPa

*Injection Area for 5 kg projectile is 94 cm², etc.

TABLE 2. Propellant Data

	Liquid JA-2	LGP 1846
Impetus (J/g)	1140	898
Flame Temp (K)	3409	2468
Gamma	1.225	1.223
Co-Volume (cm³/g)	.996	.677
c_p (J/g-K)	1.821	1.999
Density (g/cm³)	1.43	1.43

RESULTS

Computer simulations for the three test cases were made using the lumped parameter IB model developed by Coffee[11] with both the standard and modified Lagrange gradient models. Simulations were also run using an IB model developed by Gough.[2] The Gough model treats the reservoir and chamber as lumped parameter regions, and utilizes an assumption of isentropic gas expansion to connect the chamber with the barrel, which is modeled as a one-dimensional flow region. Since a one-dimensional model should provide a more accurate simulation of the barrel region, the Gough model was selected as a baseline for evaluation of the Lagrange gradient models. Both models were run assuming instantaneous burning of propellant injected into the combustion chamber, i.e. this option was used in running the simulations. As a result, the simulations of the reservoir and chamber regions are nearly identical in the Gough and Coffee models, and any differences in the overall model calculations are primarily due to differences in the simulation of the barrel region. The calculated muzzle velocities for the nine cases simulated (3 projectile masses and 3 gradient models) are presented in Table 3.

TABLE 3. Comparison of Calculated Muzzle Velocities

Projectile Mass (kg)	Gough Model One-Dimensional Flow Muzzle Velocity (m/s)	Coffee Model Standard Lagrange Gradient Muzzle Velocity (m/s)	Difference (%)	Coffee Model Modified Lagrange Gradient Muzzle Velocity (m/s)	Difference (%)
5.0	2020	2115	4.7	1952	-3.4
7.0	1894	1977	4.4	1847	-2.5
13.0	1563	1617	3.5	1546	-1.1

586

The differences between the muzzle velocities calculated using the Lagrange gradients and those obtained from the Gough models are presented as a percent of the baseline velocity. As can be seen, use of the standard Lagrange gradient consistently results in muzzle velocities higher than the baseline case, while use of the modified Lagrange gradient results in velocities which are lower than the baseline. Overall, the velocities obtained using the modified Lagrange gradient are somewhat closer to the baseline than those obtained with the standard gradient, and the agreement degrades with increasing velocity.

The pressure versus time curves from the simulations with the 7 kg projectile mass are presented in Figures 6 and 7. In Figure 6, the pressure versus time curves at three locations, chamber (top), barrel entrance (middle) and projectile base (bottom), obtained using the Coffee model with the standard Lagrange gradient (dotted lines) are presented, along with the corresponding pressure curves obtained using the Gough model for comparison (solid line). In Figure 7, the pressure versus time curves obtained using the Coffee model with the modified Lagrange gradient (dotted lines) are presented, again with results obtained using the Gough model for comparison.

As can be seen in Figure 6, the chamber pressures are quite similar, but the pressure versus time curves at the barrel entrance and projectile base show significant differences. The standard Lagrange gradient results in a pressure at the projectile base which is consistently higher than that obtained with the one-dimensional model, resulting in a higher muzzle velocity; see Table 3. The pressure at the barrel entrance for the standard Lagrange gradient case is consistently lower than that for the one-dimensional case. Overall, the standard Lagrange gradient model produces results quite different from the base line, one-dimensional model.

In comparison the modified Lagrange gradient model results show very good agreement with the one-dimensional model; see Figure 7. The
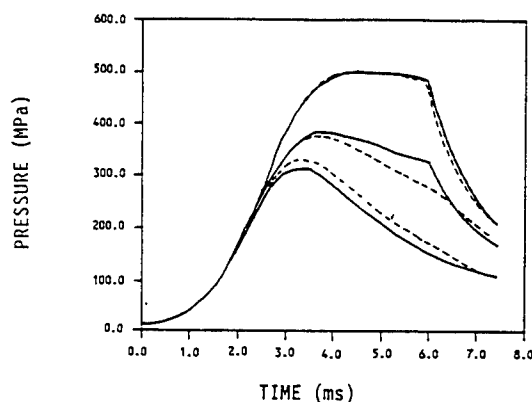

Figure 6. Comparison of Pressure vs. Time for the 7 kg Projectile; Standard Lagrange Gradient Model.
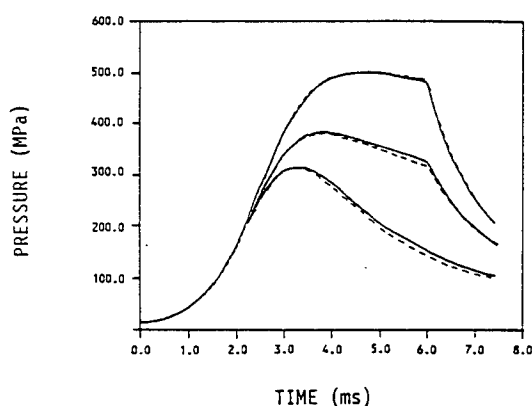

Figure 7. Comparison of Pressure vs. Time for the 7 kg Projectile; Modified Lagrange Gradient Model.

587

chamber and barrel entrance pressures are almost indistinguishable over the entire ballistic cycle for the two simulations. The projectile base pressures are also nearly indistinguishable over the first half of the ballistic cycle, but the modified Lagrange curve increasingly departs from the baseline curve after about 4.0 ms. This departure is below the baseline, consistent with the velocity difference noted in Table 3. This discrepancy between the modified Lagrange and one-dimensional results is attributed to the decreasing validity of both the basic Lagrange assumption, i.e. uniform gas density, as the projectile becomes farther removed from the barrel entrance, and the assumption of fixed v_{Lb} and $\dot{v}_{Lb}$ after all-burnt.

Pressure and velocity profiles in the barrel region from the simulations for the 7 kg projectile using the modified Lagrange and one-dimensional models are presented in Figures 8-10. Additional profiles are presented in Reference 4. The times for which these profiles were plotted were chosen to coincide with maximum chamber pressure (4.7 ms); a time when the rarefaction wave has propagated some distance down the barrel (6.5 ms); and muzzle exit (7.4 ms).
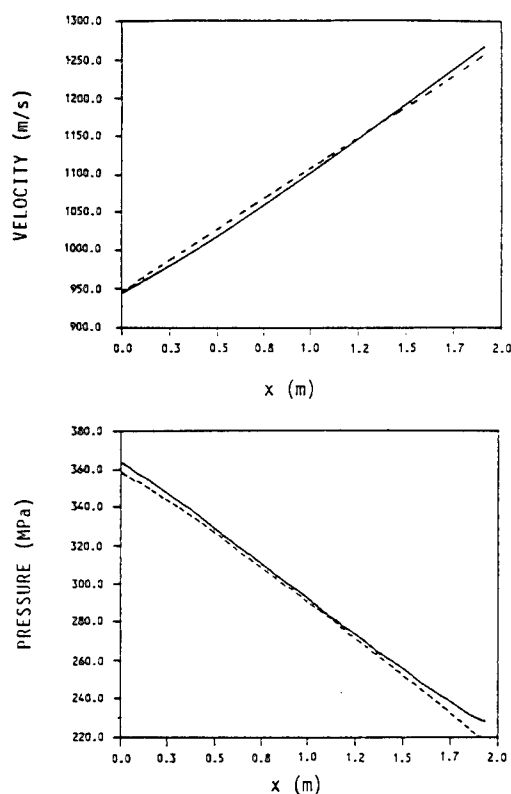
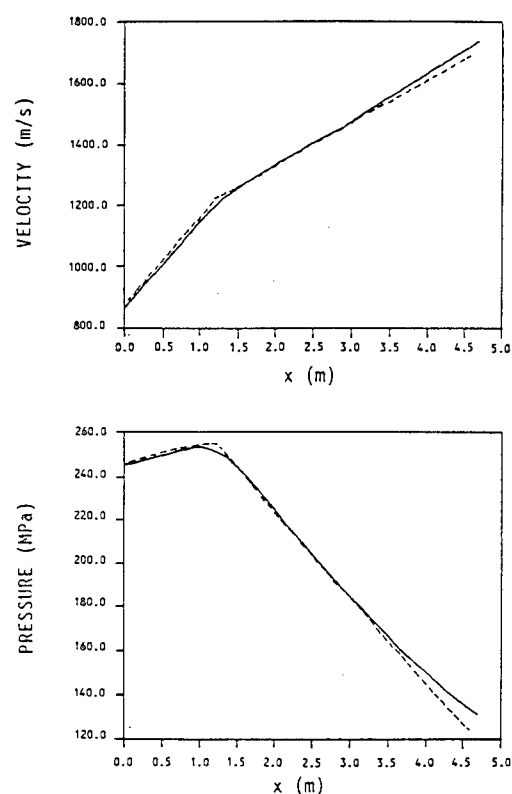Figure 8. Comparison of Velocity and Pressure Distributions in the Barrel at t = 4.7 ms.

Figure 9. Comparison of Velocity and Pressure Distributions in the Barrel at t = 6.5 ms.

Prior to all-burnt (4.7 ms) the modified Lagrange model produces velocity and pressure profiles which have the correct shape and are quite close to the one-dimensional profiles over the entire barrel region; see Figure 8. The departure in pressure at the projectile base, noted earlier, as well as a corresponding difference in velocity near the projectile base, has begun to appear.

At 6.5 ms, the rarefaction wave is well developed and has propagated about one quarter of the distance to the projectile base; see Figure 9. The velocity and pressure profiles are still in good agreement with the baseline case, and the hinge point is in approximately the correct position. The pressure and velocity differences near the projectile base have continued to increase; the pressure difference at the projectile base has increased to about 8%.

At muzzle exit, Figure 10, the velocity profiles are still in very good agreement, but the modified Lagrange pressure profile has deteriorated somewhat. The shape of the pressure profile is still approximately correct, however, the departure from the baseline profile is apparent; the pressure difference at the projectile base has increased to approximately 10%. This discrepancy does not reflect the overall accuracy of the approximation, however, since the cumulative difference in projectile velocity is only about 2.5% in this case.
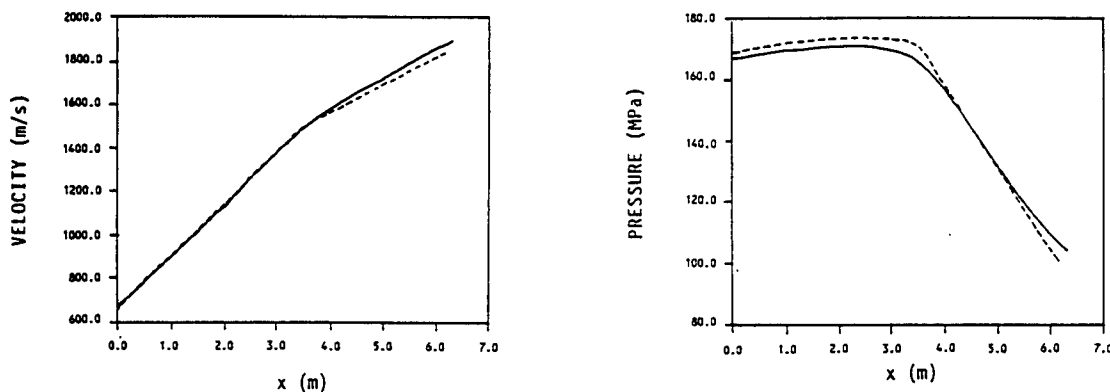

Figure 10. Comparison of Velocity and Pressure Distributions in the Barrel at t - 7.4 ms.

CONCLUSIONS

A modification of the standard Lagrange gradient model has been developed to simulate the gas flow in the barrel of a regenerative liquid propellant gun. This model accounts for the non-zero gas velocity at the entrance to the barrel, and the existence of a rarefaction wave which travels along the barrel from the chamber toward the projectile base after all-burnt. The modified Lagrange gradient has been implemented in the lumped parameter interior ballistic model developed by Coffee, and comparisons have been made among the resulting IB model, the Coffee model with a standard Lagrange

gradient, and the Gough model with a one-dimensional barrel flow model. Computer simulations of a high velocity 120-mm tank cannon have been run, using these three models, for 3 projectile masses.

The modified Lagrange gradient model resulted in muzzle velocities 1.1-3.4% lower than the one-dimensional model over the velocity range 1563-2020 m/s, while the standard Lagrange model produced muzzle velocities 3.5-4.7% higher than the one-dimensional model. The pressure versus time curves obtained using the modified Lagrange gradient are in very good agreement with those from the one-dimensional model, although some discrepancy is observed at the projectile base over the latter half of the ballistic cycle. A similar comparison for the standard Lagrange gradient case shows very poor agreement with the one-dimensional model at the barrel entrance and at the projectile base over most of the ballistic cycle.

Detailed comparisons of the velocity and pressure profiles in the barrel from the modified Lagrange and one-dimensional models at three different times during the ballistic cycle have been presented. These comparisons show that the profiles obtained using the modified Lagrange gradient have the correct shape and, with the exception of the pressure profile near muzzle exit, are in excellent agreement with the one-dimensional results. Overall, the modified Lagrange gradient model appears to be capable of accurately representing the physical processes in the barrel region of a regenerative LP gun, and of providing an excellent overall simulation of the IB process, even in the high velocity regime.

REFERENCES

1. Coffee, T.P., "A Lumped Parameter Code for Regenerative Liquid Propellant Guns," BRL Technical Report No. BRL-TR-2703, December 1985.

2. Gough, P.S., "A Model of the Interior Ballistics of Hybrid Liquid Propellant Guns," BRL Contract Report No. BRL-CR-566, March 1987.

3. Corner, J., Theory of the Interior Ballistics of Guns, J. Wiley & Sons, Inc., New York, Copyright 1950.

4. Morrison, W.F. and Coffee, T.P., "A Modified Lagrange Pressure Gradient for the Regenerative Liquid Propellant Gun," BRL Technical Report No. BRL-TR-3073, January 1990.

5. Coffee, T.P., "The Analysis of Experimental Measurements on Liquid Regenerative Guns," BRL Technical Report No. BRL-TR-2731, May 1986.

6. Gough, P.S., Contract Report BRL-CR-624, "Continuum Modeling of Regenerative Liquid Propellant Guns and Hybrid Traveling Charge Systems," December 1989.

7. Morrison, W.F., Baer, P.G., Bulman, M.J. and Mandzy, J., "The Interior Ballistics of Regenerative Liquid Propellant Guns," BRL Technical Report No. BRL-TR-2857, Oct 1987.

8. Mandzy, J., General Electric Company, private communication.

9. Gough, P.S., Contractor Report DAAK11-85-D-0002, in preparation.

10.   Baer, P.G., Coffee, T.P. and Morrison, W.F., "Design Optimization for a High Performance Regenerative Liquid Propellant Gun," BRL Technical Report No. BRL-TR-2860, October 1987.

11.   Coffee, T.P., "An Updated Lumped Parameter Code for Regenerative Liquid Propellant In-Line Guns," BRL Technical Report No. BRL-TR-2974, December 1988.

LIST OF SYMBOLS

A_B	Bore Area	P_{RES}	Barrel Resistance Pressure
a	Average Speed of Sound	T	Temperature
C	Charge Mass	u_p	Projectile Velocity
c_p	Specific Heat Constant P	v	Gas Velocity
h	Enthalpy	x	Coordinate Along the Barrel
M_p	Projectile Mass	y	Coordinate of Projectile
m	Mass of Gas in Barrel	γ	Ratio of Specific Heats
P	Pressure	ρ	Gas Density
$\bar{P}$	Space Mean Pressure		

Subscripts:

BASE	Projectile Base
BREECH	Gun Breech
H	Hinge-Point
L	Barrel Entrance
Lb	Barrel Entrance at All-Burnt
R	Projectile Base

Side Moment Exerted by a Spinning, Coning,
Highly Viscous Liquid Payload (U)

*Charles H. Murphy, Dr., James W. Bradley, Mr.
and William H. Mermagen, Sr., Mr.
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, Maryland 21005-5066

I. INTRODUCTION

Spinning projectiles containing liquid payloads have shown spectacular instabilities in flight. The first theoretical work by Stewartson[1] in 1959 showed that a spinning inviscid liquid could exert a very large side moment for certain values of τ, the ratio of coning rate to spin rate. In 1965 Wedemeyer[2] added a boundary layer on the container walls; his predicted values of the resonant side moment were verified by experiments using gyroscopes. This Stewartson-Wedemeyer (SW) boundary-layer theory was originally limited to pressure moments, fully- or partially-filled cylindrical cavities, and fully spun-up liquids. Later authors[3-8] extended the theory to include central rods, viscous moments, partially spun-up liquids, two immiscible liquids, and fully-filled spheroidal cavities. In addition, a relation between liquid side moment and liquid roll moment was derived[9] to justify experimental observation of a correlation between these moments.

In 1978 Kitchens, Gerber and Sedney[10] developed a hybrid linear Navier-Stokes theory (hereafter called KGS) that satisfied the boundary conditions on the cylindrical wall exactly by solving a sixth-order system of ordinary differential equations, but it required the addition of boundary layers on the end-walls to satisfy the no-slip condition on these walls. This hybrid high-Reynolds-number theory has been extended to predict the complete moment exerted by both fully spun-up liquids[11,12] and partially spun-up liquids.[13,14]

For very low Reynolds numbers ($Re < 100$), Vaughn et al.[15], Strikwerda et al.[16] and Rosenblat et al.[17] have developed Computational Fluid Dynamics (CFD) codes to predict moments for a fully-filled cylinder coning at a constant angle. These predictions showed good agreement with experiments, but this CFD approach requires considerable

time on large computing machines. Herbert[18,19] developed an approximate spectral method that requires rather modest computational effort and gives good agreement with the more precise CFD results for $Re < 100$.

Hall, Sedney and Gerber[20] have recently modified the hybrid Navier-Stokes theory to eliminate the approximation of the end-wall boundary conditions. This was done by replacing the simple Stewartson-Wedemeyer spatial eigenvalues by special eigenvalues whose eigenfunctions can be combined to satisfy all the boundary conditions. Once a table of eigenvalues has been constructed, this HSG method can compute liquid moments for $Re < 2500$ on VAX 8600-size computers in less than five minutes. These computations show excellent agreement with the CFD calculations[21] for $Re < 100$ and good agreement with the boundary-layer theories for $Re > 2000$.

The KGS and HSB theories are limited to a fully-filled cylindrical container performing constant-amplitude coning motion. In this paper, we will extend these theories to partially-filled cylinders and to cylinders with a central rod. We will also allow the amplitude of the motion to change slowly. In Reference [22] these theories are extended to the case of two immiscible liquids.

In the implementation of the SW, KGS and HSG theories, the liquid moment has usually been computed by integrating the pressure and wall shears over the surface of the cylindrical container. A mathematically equivalent approach is to integrate the angular momentum over the liquid volume and then differentiate this integral. It will be shown that for the essentially exact HSG theory, the two methods - surface and volume integration - yield the same value for the liquid moment. However, for the less exact SW and KGS theories, the volume integral method gives much 'better' moment values, in the sense that they are in closer agreement with HSG results. Indeed, using the volume integral, the SW theory adequately approximates HSG results down to $Re = 500$ and KGS adequately approximates HSG down to $Re = 1$.

II. EQUATIONS OF LIQUID MOTION

We will consider a projectile with a cylindrical cavity of radius a and height $2c$ and hence of fineness ratio $A = c/a$. The axis of the cylinder is collinear with the projectile's axis and the cylinder center of mass is located a distance h from the projectile's center of mass. For simplicity, we will assume that $h = 0$; the effect of a nonzero h is considered in Reference [22].

If the cavity is only partially filled, if the liquid is fully spun-up and if the centrifugal force is large compared to the aerodynamic forces, then in the absence of coning motion the liquid will fill the space between the outer cylindrical wall and an inner

594

cylindrical free surface of radius b. The ratio of the volume of this inner cylinder to the volume of the payload cavity is b^2/a^2. The fill ratio for the payload cavity is thus $1 - b^2/a^2$.

In addition to fully and partially filled cylindrical payload cavities, the case of a cylindrical cavity with a central rod of radius d can be easily treated by the linear theory. The annular region between the inner and outer cylindrical walls will be considered fully filled with liquid.

Although the differential equations of motion are most conveniently expressed in earth-fixed coordinates, the boundary conditions are best expressed in cylindrical coordinates aligned with the projectile's axis of symmetry. Symmetry-axis coordinates have one free parameter: angular velocity about the symmetry axis. If this angular velocity is chosen to be the projectile's spin rate, the system is the usual missile-fixed coordinate system. If this angular velocity is chosen to be zero, we have the nonspinning aeroballistic system used in the flight mechanics of symmetric missiles[23].

Earth-fixed axes X_e, Y_e, Z_e are selected so that X_e is initially along the velocity vector[1] and the X-axis of the missile-aligned system will be along the projectile's axis of symmetry. The angle between the X and X_e axes is the total angle of attack, α_t. The angle between the Z_e-axis and the normal to the plane formed by X and X_e will be called ϕ_α, where $\dot{\phi}_\alpha$ is the precessional rate of the angular motion performed by the projectile about the velocity vector.

Although the angular motion of a projectile does not in general have a constant precessional rate, it can be shown that the linear motion is the sum of two circular motions with constant precessional rates. The objective of all liquid payload theories is to predict the liquid moment response to constant-precession angular motion. For these theories, projectile-aligned coordinates that rotate at the precessional rate (coning aeroballistic coordinates) are especially convenient since the boundary conditions are then independent of time.

Let $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$ be unit vectors in the earth-fixed system and let $(\hat{e}_{xc}, \hat{e}_{yc}, \hat{e}_{zc})$ be unit vectors in the coning aeroballistic system. The angular velocity $\vec{\Omega}$ of the coning aeroballistic coordinate system is defined by the relation

$$(\dot{\hat{e}}_{xc}, \dot{\hat{e}}_{yc}, \dot{\hat{e}}_{zc}) = \vec{\Omega} \times (\hat{e}_{xc}, \hat{e}_{yc}, \hat{e}_{zc}) \tag{1}$$

It can be easily verified that

$$\vec{\Omega} = \tau\dot{\phi}[\hat{e}_x - (\epsilon K/\gamma)\hat{e}_{zc}] \tag{2}$$
$$\hat{e}_x = \gamma\hat{e}_{xc} + K\hat{e}_{yc} \tag{3}$$

---

[1]In this paper, we assume that the velocity vector maintains a constant direction. The effects of aerodynamic forces and gravity are given in Reference [23].

where

- $\dot{\phi}$ is the projectile spin rate relative to an inertial frame;

- $\tau = \dot{\phi}_\alpha/\dot{\phi}$, the ratio of the precessional rate to the spin rate; and

- $K = K_0 e^{\epsilon\tau\phi}$. Hence $\epsilon\tau\dot{\phi} = \dot{K}/K$ and we see that ϵ is a measure of the logarithmic damping during a precession cycle. Zero ϵ denotes constant-amplitude coning motion.

Cylindrical coordinates will be denoted by (x, r, θ) in the earth-fixed system and by $(\tilde{x}, \tilde{r}, \tilde{\psi})$ in the coning aeroballistic system. (For convenience, we will assume that all length variables have been made dimensionless by division by the radius a.) The earth-fixed cylindrical components of the velocity of a point on the projectile are

$$V_x = a\dot{\phi}R\{i(f-1)rK^*\} \tag{4}$$
$$V_r = -a\dot{\phi}R\{i(f-1)xK^*\} \tag{5}$$
$$V_\theta = a\dot{\phi}[r - R\{(f-1)xK^*\}] \tag{6}$$

where

$$K^* = Ke^{-i\psi} = K_0 e^{i(f\phi-\theta)}$$
$$f = (1 - i\epsilon)\tau$$
$$\psi = \theta - \phi_\alpha = \theta - \tau\phi$$
$$\phi = \dot{\phi}t$$
$$\phi_\alpha = \dot{\phi}_\alpha t$$

and where $R\{\ \}$ denotes the real part of a complex quantity.

We will now make the very restrictive assumption that the liquid is in steady-state response to the coning and spinning motion of the projectile. Theoretical studies[5,6] have been made and are in progress to determine the effect of partially spun-up liquid, and an experimental study[24] has been made of the transient response to coning motion. These studies show that spin-up and cone-up effects are large and important to a complete understanding of the liquid-payload stability problem.

Nevertheless, we will assume that the liquid velocity components and liquid pressure have the same dependency on time and θ as do the velocity components of points on the projectile. Accordingly, we introduce three dimensionless complex perturbation velocities: $\underline{w}, \underline{u}, \underline{v}$, each a function of r and x, and write:

$$V_x = -a\dot{\phi}R\{\underline{w}K^*\} \tag{7}$$
$$V_r = -a\dot{\phi}R\{\underline{u}K^*\} \tag{8}$$
$$V_\theta = a\dot{\phi}[r - R\{\underline{v}K^*\}] \tag{9}$$

The equilibrium pressure for a spinning non-coning cylinder is

$$p_{eq} = p_0 + \rho_L a^2 \dot{\phi}^2 \left[\frac{r^2 - r_0^2}{2} \right] \quad (r_0 \leq r \leq 1) \tag{10}$$

where r_0 is either b/a or d/a and where ρ_L is the liquid density.

Each of the four perturbation functions can be assumed to be the sum of products of two variables: the first a function of r alone and the second a function of x alone:

$$\underline{w} = i(1-f)rH_0(x) - f\sum_{k=0}^{N} c_k w_k(r)H_k(x) \tag{11}$$

$$\underline{u} = -\left[\frac{i(1-f)^2}{1+f} \right] G_0(x) + f\sum_{k=0}^{N} c_k u_k(r)G_k(x) \tag{12}$$

$$\underline{v} = -\left[\frac{(1-f)^2}{1+f} \right] G_0(x) + f\sum_{k=0}^{N} c_k v_k(r)G_k(x) \tag{13}$$

$$\underline{p} = -(1-f)^2 rG_0(x) + f\sum_{k=0}^{N} c_k p_k(r)G_k(x) \tag{14}$$

where

$$for \ \lambda_0 = 0, \ H_0 = 1 \ and \ G_0 = x \tag{15}$$

$$for \ \lambda_k \neq 0, \ H_k = \cos(\lambda_k x)/\sin(\lambda_k A) \ and \ G_k = \sin(\lambda_k x)/\sin(\lambda_k A) \tag{16}$$

The λ_k's are complex parameters arising from this classical separation-of-variables technique and will be evaluated so as to help satisfy the boundary conditions. (The values so determined will be referred to as eigenvalues.) The coefficients of H_0 and G_0 in Eqs.(11-14) are special perturbation functions for $\lambda_0 = 0$. In the low-viscosity theories of References [4, 10-12]:

$$w_0 = u_0 = v_0 = p_0 = 0 \tag{17}$$

Eqs.(11-14) can be substituted in the linearized Navier-Stokes equations to obtain

$$ru'_k + u_k - iv_k + \lambda_{k1}rw_k = 0 \tag{18}$$

$$(Re)^{-1}[u''_k + u'_k/r - (B + r^{-2})u_k] + 2[1 + i(r^2 Re)^{-1}]v_k = p'_k \tag{19}$$

$$(Re)^{-1}[v''_k + v'_k/r - (B + r^{-2})v_k] - 2[1 + i(r^2 Re)^{-1}]u_k = -ip_k/r \tag{20}$$

$$(Re)^{-1}[w''_k + w'_k/r - Bw_k] = -\lambda_{k2}p_k \tag{21}$$

where

$$B = r^{-2} + \lambda_k^2 - i(1-f)Re$$
$$\lambda_{01} = 0, \ \lambda_{02} = 1$$
$$\lambda_{k1} = \lambda_{k2} = \lambda_k, \ k \neq 0$$
$$(\ )' = d(\ )/dr$$

III. LOW-VISCOSITY SOLUTIONS

For high Reynolds numbers ($Re > 5000$), three distinct solutions have been obtained to the preceding differential equations: (1) Stewartson (S) inviscid solution; (2) Stewartson-Wedemeyer (SW) inviscid solution modified by wall boundary layers; (3) Kitchens-Gerber-Sedney (KGS) linear Navier-Stokes solution with end-wall boundary layers. In all three methods, the flat end-wall boundary condition on the liquid normal velocity is satisfied by omitting half the eigenfunctions (i.e., $c_k = 0$ for k even) and by a proper selection of the remaining eigenvalues λ_k.

(1) <u>Stewartson (S) Solution</u>. For infinite Reynolds number, Eqs.(18-21) can be manipulated to yield a second-order differential equation for the p_k's and algebraic conditions for the perturbation velocities u_k, v_k, w_k in terms of p_k and p'_k. The appropriate eigenvalues have the very simple form

$$\lambda_k = \frac{k\pi}{2A} \tag{22}$$

The radial boundary conditions are set by the normal velocity at the outer cylindrical wall and different inner conditions for the three cases of fully filled, partially filled and inner cylindrical rod. These can be satisfied by simple boundary conditions on the p_k's and the remaining c_k's.

(2) <u>Stewartson-Wedemeyer (SW) Solution</u>. The Stewartson inviscid solution could not satisfy the tangential velocity conditions on the container walls. In 1965 Wedemeyer introduced three oscillatory boundary layers on these walls. These boundary layers created small viscous contributions to the liquid moment, but more importantly they modified the larger inviscid pressure moment term. The boundary layers caused an outflow velocity that changed the normal velocity conditions for the inviscid solution. The presence of the boundary layers on the end walls modifies the eigenvalues, λ_k, and the Fourier coefficients, c_k. The new eigenvalues can be easily computed by the relation

$$\lambda_k = \frac{k\pi}{2(A - \delta_c)} \tag{23}$$

where

$$\delta_c = \frac{-(1+i)}{2(1-f)\sqrt{2Re}} \left[\frac{1+f}{\sqrt{3-f}} + \frac{i(3-f)}{\sqrt{1+f}} \right]$$

3. <u>Kitchens-Gerber-Sedney (KGS) Solution</u>. The KGS solution employs the full linearized Navier-Stokes equations but retains the Wedemeyer end-wall boundary layers. This allows the use of the easily computed SW eigenvalues of Eq.(23). The

cylindrical boundary layer is eliminated and the usual no-slip conditions are satisfied exactly.

The KGS solution is probably better than the SW solution since it depends on fewer boundary-layer approximations; however, KGS has been limited to applications for which $Re > 1000$. An important characteristic of the KGS method is the fact that it can be modified to give an exact low-Reynolds-number solution (the HSG solution).

IV. HIGH-VISCOSITY (HSG) SOLUTION

In the low-viscosity solutions, the $k = 0$ solution was selected to satisfy the boundary conditions at the flat end-walls. The λ_k's were then the eigenvalues of the differential equations for functions of x. The boundary conditions on the cylindrical wall were then satisfied by a Fourier series fit of these functions.

The HSG method reverses this process. The $k = 0$ solution is selected to satisfy the boundary conditions on the cylindrical surface. The λ_k's become the eigenvalues of the differential equations for functions of r and the end-wall boundary conditions are satisfied by a least-squares fit to eigenfunctions of r.

The actual determination of the eigenfunctions is a delicate process. A simple Newton iterative method (strictly speaking, the *regula falsi* method) has been coded; it will converge rapidly to an eigenvalue *if* a good initial guess is available. For fully filled cylinders, look-up eigenvalue tables have been constructed for Re between 1 and 2500 and for τ between -1 and 1.5. Less complete tables are available for the partial fill and central rod cases. Hall et al. showed[20] that for a fully-filled cylinder, each eigenvalue falls into one of three families. Figure 1 (from Reference [20]) shows 72 of these eigenvalues for $Re = 1000$, $\tau = 0.1$. (Note that eigenvalue (3,1) is a maverick in this three-family concept.)

V. LIQUID MOMENT

If a projectile is performing simple coning motion, the moment exerted on it by a fully spun-up liquid payload can be defined in terms of three dimensionless moment coefficients. These coefficients are related to the liquid moment (M_{xc}, M_{yc}, M_{zc}) in the coning coordinate system as follows:

$$M_{xc} = m_L a^2 \dot{\phi}^2 \tau C_{LRM} K^2 \tag{24}$$

$$M_{yc} + i M_{zc} = m_L a^2 \dot{\phi}^2 \tau [C_{LSM} + i C_{LIM}] K \tag{25}$$

where

$$m_L = 2\pi a^2 c \rho_L$$

These three coefficients - the liquid roll moment coefficient C_{LRM}, the liquid side moment coefficient C_{LSM}, and the liquid in-plane moment coefficient C_{LIM} - are all functions of A, τ, ϵ, Re and K^2. The liquid side moment affects the damping of the coning motion while the liquid in-plane moment can cause a frequency shift. Thus our primary interest is in the liquid side moment coefficient C_{LSM} since it determines any instabilities that could occur.

The linear coefficients can be computed by integrating the pressures and viscous stresses over the lateral cylindrical wall and the two flat end-walls. If a central rod is present, its pressure and viscous contributions must also be included. Using subscripts p for pressure, v for viscous, ℓ for lateral wall, e for end wall and r for rod, we have[4,12]

$$C_{LSM} + iC_{LIM} = C_{mp\ell} + C_{mpe} + C_{mpr} + C_{mv\ell} + C_{mve} + C_{mvr} \tag{26}$$

Expressions for the components in Eq.(26) are given in Ref.[22].

The linear fluid moments can also be computed directly by differentiating the volume integral of the angular momentum of all the liquid in the cylinder. To do this, we first express the position and velocity vectors in earth-fixed cylindrical coordinates $(\hat{e}_x, \hat{e}_r, \hat{e}_\theta)$:

$$\vec{R} = (x\hat{e}_x + r\hat{e}_r)a \tag{27}$$

$$\vec{V} = V_x\hat{e}_x + V_r\hat{e}_r + V_\theta\hat{e}_\theta \tag{28}$$

Then

$$\vec{R} \times \vec{V} = [rV_\theta\hat{e}_x - xV_\theta\hat{e}_r + (xV_r - rV_x)\hat{e}_\theta]a \tag{29}$$

The angular momentum is

$$\begin{aligned}
\vec{L} &= a^3 \int_0^1 \int_{-A}^A \int_0^{2\pi} \rho \vec{R} \times \vec{V}\tilde{r}\, d\tilde{r}\, d\tilde{x}\, d\tilde{\psi} \\
&= L_1\hat{e}_{xc} + L_2\hat{e}_{yc} + L_3\hat{e}_{zc} \tag{30}
\end{aligned}$$

The moment exerted by the liquid on the projectile is the negative of the derivative of the liquid's angular momentum:

$$\begin{aligned}
\vec{M} &= -\dot{\vec{L}} \\
&= -(\dot{L}_1\hat{e}_{xc} + \dot{L}_2\hat{e}_{yc} + \dot{L}_3\hat{e}_{zc} + \vec{\Omega} \times \vec{L}) \tag{31}
\end{aligned}$$

For steady-state constant-amplitude liquid motion,

$$\dot{L}_j = \epsilon = 0 \tag{32}$$

$$\begin{aligned}
\vec{M} &= -\tau\dot{\phi}\hat{e}_x \times \vec{L} \\
&= \tau\dot{\phi}[-KL_3\hat{e}_{xc} + \gamma L_3\hat{e}_{yc} + (KL_1 - \gamma L_2)\hat{e}_{zc}] \tag{33}
\end{aligned}$$

Eq.(33) - applied to definitions (24-25) - yields the very important result that

$$\gamma C_{LRM} = -C_{LSM} \tag{34}$$

This result has been obtained elsewhere[9,17], but the above is the simplest, most direct derivation and is based solely on the assumption of steady-state, constant-amplitude motion given by Eq.(32).

VI. NUMERICAL RESULTS

(1) Surface Integral versus Volume Integral. In the HSG, KGS and SW theories, the liquid moments were originally computed by integrating the pressure and wall shears over the walls of the container. The volume integral method should give the same results if the velocity perturbation functions satisfy the linearized Navier-Stokes equations and the boundary conditions. For a sufficient number of eigenfunctions, this is the case for the HSG theory: the two values of C_{LSM} are equal. In Figure 2, C_{LSM} values obtained from HSG by both the surface and volume integral methods are plotted versus the number of eigenvalues used, for $Re = 500$, $A = 3.1$ and $\tau = 0.1$. For 28 eigenvalues or more, the values from the two methods are essentially equal. For a smaller number of eigenvalues, the two values differ, but the volume integral result is always closer to their common value for more than 28 eigenvalues. This figure also shows the error of the fit to the end-wall boundary condition; we see that the C_{LSM} are the same when the error is less than 0.10.

Gans[25] uses the volume integral method in his version of SW and he notes that this method correctly predicts that the side moment coefficient goes to zero as τ goes to zero, whereas SW results based on the surface integral method do not have this property. Both Rosenblat[17] in his finite element calculations and Herbert[18] in his spectral method use the volume integral approach. Herbert emphasizes the value of the volume integral in giving better results for the approximate theories. Indeed, for the approximate boundary layer theories at low Reynolds numbers, the average error in velocity predictions should be much less than the error in the velocity gradients at the boundary.

In Figure 3, the volume and surface integral results for KGS and SW are compared at $Re = 500$, $A = 3$ and τ ranging from 0 to 0.20. The exact HSG values are also given in the figure. The surface integral results are essentially worthless, but the volume integral results are quite good, with KGS somewhat better than SW. For this reason, we will routinely use the volume integral to compute the liquid moments for all theories.

(2) Fully Filled Cylinders. Historically, our interest in the moment produced by

highly viscous liquids is based on experiments by Miles Miller[26]. He measured the liquid roll moment in a coning, spinning, liquid-filled cylinder and showed that at a Reynolds number of about 150, the liquid roll moment was large enough to cause flight instabilities[27]. The theoretical basis for this very important result is Eq.(34), which directly relates the liquid roll moment of Miller's experiments to the liquid side moment that controls flight instabilities.

In Figure 4, we plot the negative of Miller's[28,29] C_{LRM} versus $log_{10}Re$ for a fully filled cylinder of fineness ratio 4.5 and $\tau = 0.1$. This figure also gives the HSG, KGS and SW curves for C_{LSM}. For HSG and KGS, we see good qualitative agreement but poor quantitative agreement with the Miller data for Re less than about 20.

The most striking feature revealed by Figure 4 is the exceptional ability of the KGS theory to predict the much more soundly based HSG curve. The three theories are compared for two other fineness ratios, 3 and 1.5, in Figures 5 and 6, respectively, and we see that the excellent agreement for the KGS theory persists for all Reynolds numbers greater than unity. The SW theory, however, does surprisingly well down to an Re of 100. Previous comparisons showed poor agreement for Re less than 1000. The primary difference between the earlier computations and those of Figures 4-6 is that Figures 4-6 were obtained by (a) using the volume integral method, (b) adding a special eigenvalue[30], and (c) imposing a more exact boundary condition.

The SW method results in a very fast computer code since it is based on linear combinations of easily generated Bessel functions. Indeed, three-dimensional surfaces of C_{LSM} as functions of any *two* of the parameters $(\tau, \epsilon, Re, A, r_0)$ can be obtained quickly in the interactive mode on our VAX 8600.

The KGS method solves a sixth-order complex differential equation numerically for a set of eigenvalues. These eigenvalues are quite easy to compute; usually, less than twelve are required to satisfy the boundary conditions adequately. For Re less than 1000, a single calculation of C_{LSM} takes less than 15 seconds on our VAX, but the run time increases with Reynolds number. The very exact HSG method unfortunately requires many more eigenvalues (our code has an arbitrary, but easily extendable, limit of 72). These eigenvalues are computed from the tables of starting estimates and our iterative Newton scheme.

The agreement between the theories as demonstrated by Figures 4-6 has a very important pay-off. For Reynolds numbers greater than 300, the improved SW method of this paper can be used to compute C_{LSM} in a very fast interactive code. For Re between 0 and 300, HSG results can be reasonably well approximated by the KGS code, a process somewhat slower than the SW program but still fully interactive.

This ability to obtain C_{LSM} rapidly over a wide spectrum of Re values is demon-

strated by Figure 7, a three-dimensional plot of C_{LSM}. versus Re and τ, for a fully filled cylinder of fineness ratio 4.5. The plot was obtained from the KGS and SW codes. For the higher Reynolds numbers, we see the characteristic resonance peak at about $\tau = 0.20$. For a fixed τ, a mild peak appears near $Re = 50$. Similar 3D figures were first made by Miller[28].

(3) Partially Filled Cylinders and Central Rods. Stewartson's inviscid work[1] and Wedemeyer's viscous extension[2] covered fully and partially filled cylinders. Frasier[3] extended the SW theory to a cylindrical cavity with a central rod. Both the KGS and HSG theories[10–13,20] were originally restricted to fully filled cylindrical cavities. It is quite easy to extend the KGS theory to partially filled cavities and central rods; this has been done for this paper. As mentioned earlier, the extension of the HSG theory to these cases requires the forming of additional files of eigenvalues. These eigenvalues have been obtained, using BRL's CRAY supercomputers.

For Figure 8, the side moment coefficient has been computed by all three theories for a partially filled cylinder of fineness ratio 3, with a Reynolds number of 500 and $\tau = 0.10$, and plotted against the inner cylindrical free surface ratio b/a. As we would expect, the side moment vanishes when this ratio is 1.0 and the cavity is empty. The generation of HSG eigenfunctions gets to be too tedious for $b/a > 0.9$, but we see that the KGS theory is a good approximation to the more exact HSG theory in the range $0 < b/a < 0.9$.

Figure 9 is a similar plot for the central rod ratio d/a varying from 0 to 1. The HSG eigenfunctions were only calculated for d/a up to 0.5. Once again, the computationally simpler KGS theory is a good approximation to the HSG theory. Thus for most engineering calculations, the KGS theory would be the one to use.

VII. SUMMARY

1. The four liquid payload theories - S, SW, KGS, and HSG - have been developed here in a single theoretical framework.

2. A very simple and elegant derivation of the relation between liquid side moment and liquid roll moment is given.

3. Moment coefficients have been computed by both surface integral and volume integral and the clear superiority of the volume integral is demonstrated.

4. Eigenvalue tables for a wide range of τ's and Reynolds numbers have been generated to allow quicker HSG calculation of the side moment coefficient. These

tables originally applied only to a fully filled cylindrical cavity but they have been expanded to include partially filled cylinders and cylinders with a central rod.

5. Side moment coefficients computed from the more convenient KGS theory are shown to be good approximations to those obtained by the more exact HSG theory. This agreement is shown to exist not only for fully filled cylinders but for partially filled cylinders and cylinders with a central rod as well.

References

[1] Stewartson, K., "On the Stability of a Spinning Top Containing Liquid," *Journal of Fluid Mechanics*, Vol. 5, Part 4, September 1959, pp. 577-592.

[2] Wedemeyer, E.H., "Viscous Corrections to Stewartson's Stability Criterion," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Report BRL-R-1325, June 1966. (AD 489687)

[3] Frasier, J.T., "Dynamics of a Liquid Filled Shell: Viscous Effects in a Cylindrical Cavity with a Central Rod," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Memorandum Report BRL-MR-1959, January 1969. (AD 684344)

[4] Murphy, C.H., "Angular Motion of a Spinning Projectile with a Viscous Liquid Payload," *Journal of Guidance, Control, and Dynamics*, Vol. 6, July-August 1983, pp. 280-286. [See also US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Memorandum Report ARBRL-MR-03194, August 1982. (AD A118676)]

[5] Sedney, R. and Gerber, N., "Viscous Effects in the Wedemeyer Model of Spin-Up from Rest," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02493, June 1983. (AD A129506)

[6] Murphy, C.H., "Moment Induced by a Liquid Payload During Spin-Up Without a Critical Layer," *Journal of Guidance, Control, and Dynamics*, Vol. 8, May-June 1985. [See also US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02581, August 1984. (AD A145716)]

[7] Murphy, C.H., "Side Moment Exerted by a Two-Component Liquid Payload on a Spinning Projectile," *Journal of Guidance, Control, and Dynamics*, Vol. 10, January-February 1987. [See also US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02624, December 1984. (AD A149845)]

[8] Murphy, C.H., "Moment Exerted on a Coning Projectile by a Spinning Liquid in a Spheroidal Cavity," *AIAA Journal*, Vol. 6, December 1987, pp. 1631-1633. [See also US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02755, December 1986. (AD A177642)]

[9] Murphy, C.H., "Liquid Payload Roll Moment Induced by a Spinning and Coning Projectile," *AIAA* Paper No. 83-2142, *AIAA* Atmospheric Flight Mechanics Conference, August 1983. [See also US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02521, September 1983. (AD 133681)]

[10] Kitchens, C.W., Jr., Gerber, N., and Sedney, R., "Oscillations of a Liquid in a Rotating Cylinder: Part I. Solid Body Rotation," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02081, June 1978. (AD A057759)

[11] Gerber, N., Sedney, R., and Bartos, J., "Pressure Moment on a Liquid-Filled Projectile: Solid Body Rotation," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02422, October 1982. (AD A120567)

[12] Gerber, N. and Sedney, R., "Moment on a Liquid-Filled Spinning and Nutating Projectile: Solid Body Rotation," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02470, February 1983. (AD A125332)

[13] Gerber, N., "Contribution of Pressure to the Moment During Spin-Up on a Nutating Liquid-Filled Cylinder: Ad Hoc Model," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02563, June 1984. (AD A143678)]

[14] Gerber, N., "Liquid Moment on a Filled Coning Cylinder During Spin-Up: Ad Hoc Model," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report ARBRL-TR-02628, December 1984. (AD A150280)

[15] Vaughn, H.R., Oberkampf, W., and Wolfe, W.R., "Fluid Motion Inside a Spinning Nutating Cylinder," *Journal of Fluid Mechanics*, Vol. 150, 1985, pp. 121-138.

[16] Strikwerda, J.C. and Nagel, Y.M., "A Numerical Method for the Incompressible Navier-Stokes Equations in Three-Dimensional Cylindrical Geometry," *Journal of Computational Physics* 78, 1988, pp.64-78.

[17] Rosenblat, S., Gooding, A. and Engleman, M.S., "Finite Element Calculations of Viscoelastic Fluid Flow in a Spinning and Nutating Cylinder," U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, Report CRDEC-CR-87021, December 1986.

[18] Herbert, Thorwald, "Viscous Fluid Motion in a Spinning and Nutating Cylinder," Chemical Research and Development Center, Aberdeen Proving Ground, MD, Report CRDC-CR-86004, November 1985.

[19] Herbert, Th., and Li, R., "Numerical Study of the Flow in a Spinning and Nutating Cylinder," *AIAA* Paper No. 87-1445, *AIAA* 19th Fluid Dynamics, Plasma Dynamics and Laser Conference, June 1987.

[20] Hall, P., Sedney, R., and Gerber, N., "Fluid Motion in a Spinning, Coning Cylinder via Spatial Eigenfunction Expansion," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report BRL-TR-02813, August 1987. (AD A190758)

[21] Nusca, M.J., "Computational Fluid Dynamics Methods for Low Reynolds Number Precessing/Spinning Incompressible Flows," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Memorandum Report BRL-MR-03657, April 1988. (AD A193891)

[22] Murphy, Charles H., Bradley, James W., and Mermagen, William H., Sr., "Side Moment Exerted by a Spinning, Coning, Highly Viscous Liquid Payload," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Report BRL-TR-3074, December 1989.

[23] Murphy, C.H., "Free Flight Motion of Symmetric Missiles," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Report R1216, July 1963. (AD 442757)

[24] D'Amico, W.P., Beims, W.G., and Rogers, T.H., "Pressure Measurements of a Rotating Liquid for Impulsive Coning Motion," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Memorandum Report BRL-MR-03208. (AD 121603) (See also *AIAA* Paper 82-0249, January 1982.)

[25] Gans, Roger F., "Dynamics of a Near-Resonant Fluid-Filled Gyroscope," *AIAA Journal*, Vol.22, No.10, October 1984, pp.1465-1471.

[26] Miller, Miles C., "Flight Instabilities of Spinning Projectiles Having Non-Rigid Payloads," *Journal of Guidance, Control and Dynamics*, Vol. 5, March-April 1982, pp.151-157.

[27] D'Amico, W.P. and Miller, M.C., "Flight Instability Produced by a Rapidly Spinning, Highly Viscous Liquid," *Journal of Spacecraft and Rockets*, Vol. 16, January-February 1979, pp.62-64.

[28] Miller, Miles C., "Liquid Filled Projectiles - New problems, New Solutions," *Proceedings of the 9th International Symposium on Ballistics*, American Defense Preparedness Association, Shrivenham, England, 29-30 April, 1 May 1986, pp. 2-67 to 2-76.

[29] Miller, Miles C., "Laboratory Test Fixture for Non-Rigid Payloads," *Proceedings of the 13th International Congress on Instrumentation in Aerospace Simulation Facilities* Göttingen, FRG, 18-21 September 1989.

[30] Sedney, Raymond and Gerber, Nathan, "Asymptotic Analysis of the Roots of a Certain Transcendental Equation," US Army Ballistic Research Laboratory (Aberdeen Proving Ground, MD) Technical Report TR-2727, April 1986. (AD 168315)
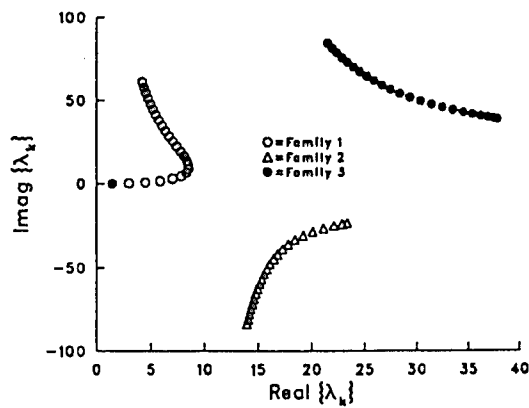
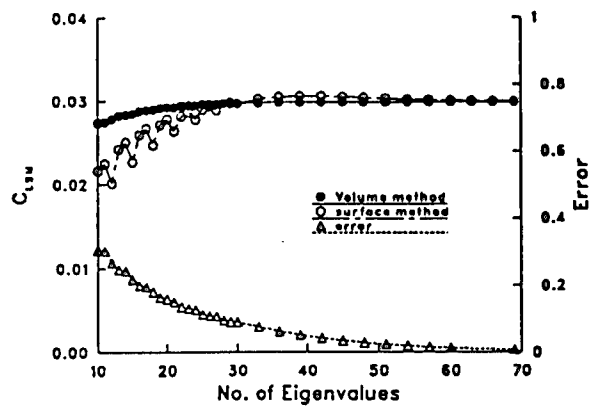Fig. 1. Eigenvalues for a fully filled cylinder,
Re = 1000, $\tau = 0.1$

Fig.2 C_{LSM} by HSG, volume and surface methods,
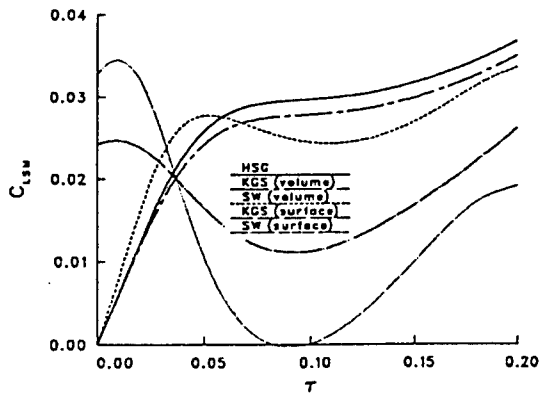vs. no. of e.v.'s (Re = 500, A = 3, $\tau = 0.1$)

Fig. 3. C_{LSM} (by KGS and SW, volume and surface
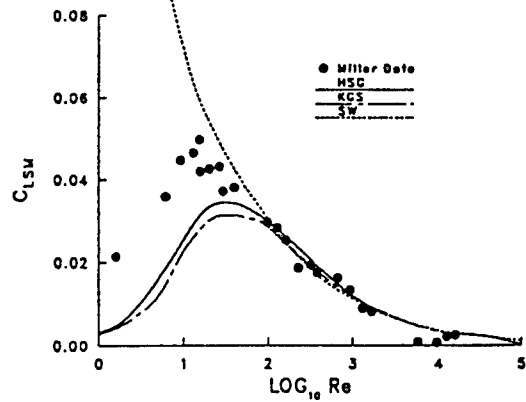methods) vs. τ (Re = 500, A = 3).

Fig. 4. C_{LSM} vs. $\log_{10}$ Re
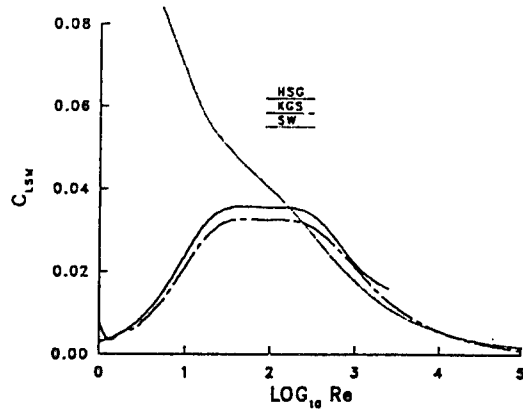for A = 4.5, $\tau = 0.1$. Comparison with Miller data.

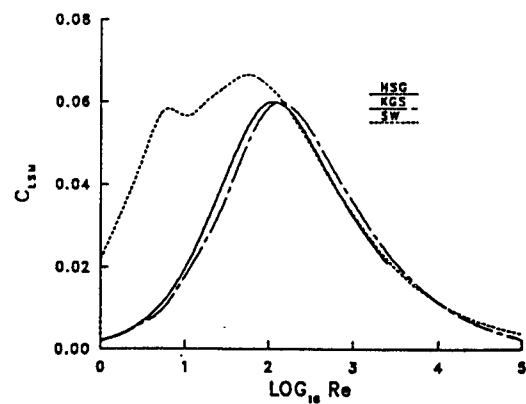Fig. 5 C_{LSM} vs. $\log_{10}$ Re
for A = 3, $\tau = 0.1$

Fig. 6. C_{LSM} vs. $\log_{10}$ Re   for A = 1.5, $\tau = 0.1$
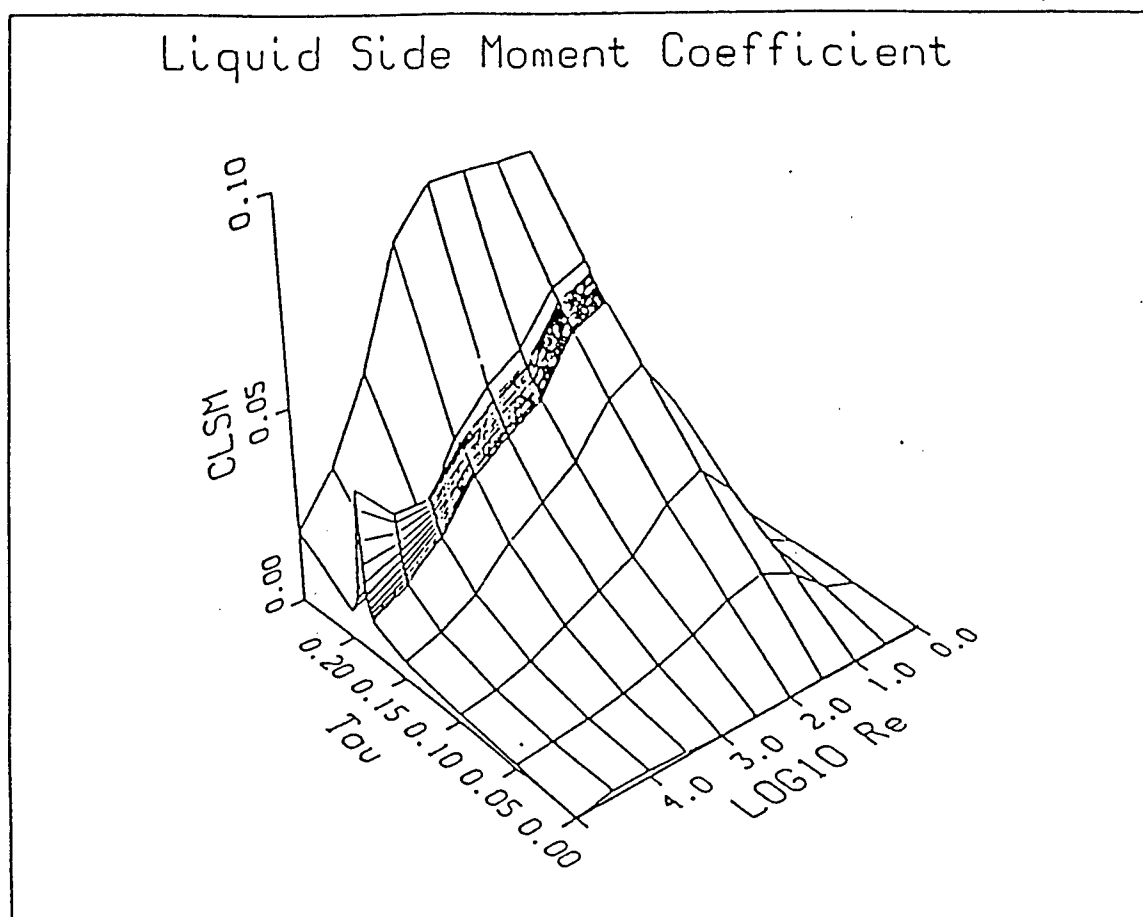
Liquid Side Moment Coefficient
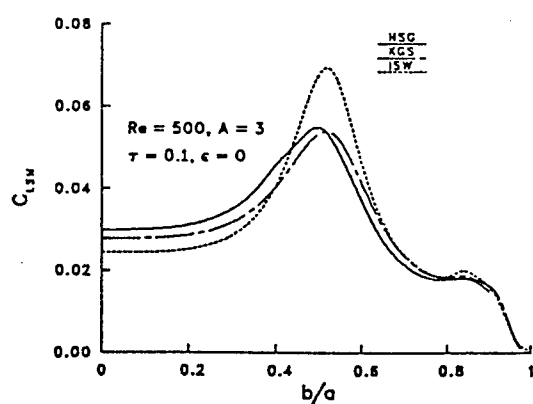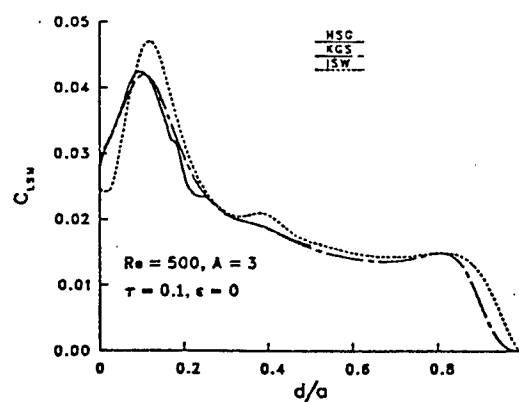
Figure 7.   3D C_{LSM} surface vs Re and τ .

Fig. 8 Partial Fill: C_{LSM} vs. b/a

Fig. 9 Central Rod: C_{LSM} vs. d/a